



**AÇIK UÇLU VE ÇOKTAN SEÇMELİ MADDELERDEN OLUŞAN
MADDE TAKIMLARININ GÜVENİRLİĞE ETKİSİNİN
GENELLENEBİLİRLİK KURAMI İLE İNCELENMESİ**

Serpil Kocaoğlu

YÜKSEK LİSANS TEZİ

EĞİTİM BİLİMLERİ ANABİLİM DALI

EĞİTİMDE ÖLÇME VE DEĞERLENDİRME BİLİM DALI

GAZİ ÜNİVERSİTESİ

EĞİTİM BİLİMLERİ ENSTİTÜSÜ

EYLÜL, 2023

TELİF HAKKI VE TEZ FOTOKOPİ İZİN FORMU

Bu tezin tüm hakları saklıdır. Kaynak göstermek koşuluyla tezin teslim tarihinden itibaren(....) ay sonra tezden fotokopi çekilebilir.

YAZARIN

Adı : Serpil
Soyadı : KOCAOĞLU
Bölümü : Eğitimde Ölçme ve Değerlendirme
İmza :
Teslim tarihi :

TEZİN

Türkçe Adı : Açık Uçlu ve Çoktan Seçmeli Maddelerden Oluşan Madde Takımlarının Güvenirliliğe Etkisinin Genellenebilirlik Kuramı ile İncelenmesi

İngilizce Adı : Investigation of the Effect of Testlets Consisting of Open-Ended and Multiple Choice Items on Reliability via the Generalizability Theory

ETİK İLKELERE UYGUNLUK BEYANI

Tez yazma sürecinde bilimsel ve etik ilkelere uyduğumu, yararlandığım tüm kaynakları kaynak gösterme ilkelerine uygun olarak kaynakçada belirttiğimi ve bu bölümler dışındaki tüm ifadelerin şahsıma ait olduğunu beyan ederim.

Yazar Adı Soyadı: Serpil KOCAOĞLU

İmza:

JÜRİ ONAY SAYFASI

Serpil KOCAOĞLU tarafından hazırlanan “Açık Uçlu ve Çoktan Seçmeli Maddelerden Oluşan Madde Takımlarının Güvenirliğe Etkisinin Genellenebilirlik Kuramı ile İncelenmesi” adlı tez çalışması aşağıdaki jüri tarafından oy birliği / oy çokluğu ile Gazi Üniversitesi Eğitim Bilimleri Anabilim Dalı, Eğitimde Ölçme ve Değerlendirme Bilim Dalı’nda Yüksek Lisans tezi olarak kabul edilmiştir.

Danışman: Doç. Dr. Melek Gülşah ŞAHİN

(Eğitimde Ölçme ve Değerlendirme, Gazi Üniversitesi)

Başkan: Prof. Dr. İsmail KARAKAYA

(Eğitimde Ölçme ve Değerlendirme, Gazi Üniversitesi)

Üye: Doç. Dr. Gülşen TAŞDELEN TEKER

(Tıp Eğitimi ve Bilişimi, Hacettepe Üniversitesi)

Tez Savunma Tarihi: 08/08/2023

Bu tezin Eğitim Bilimleri Anabilim Dalı’nda Yüksek Lisans tezi olması için şartları yerine getirdiğini onaylıyorum.

Prof. Dr. Şaban Çetin

Eğitim Bilimleri Enstitü Müdürü



Kızıma

TEŞEKKÜR

Bu tezin tamamlanmasında bana rehberlik eden, yardımlarını ve desteğini benden esirgemeyen tez danışmanım ve değerli hocam Sayın Doç. Dr. Melek Gülşah ŞAHİN'e teşekkürlerimi ve şükranlarımı sunarım.

Tez jürimde bulunarak çok önemli katkılar sunan, değerli zamanlarını bana ayırarak yol gösteren değerli hocalarım Sayın Prof. Dr. İsmail KARAKAYA ile Sayın Doç. Dr. Gülşen TAŞDELEN TEKER'e teşekkürü bir borç bilirim.

Yüksek lisans eğitimimiz boyunca değerli bilgi ve deneyimlerini bizlerle paylaşan Gazi Üniversitesi Eğitimde Ölçme ve Değerlendirme Bilim Dalındaki hocalarıma şükranlarımı sunarım.

Uygulamalarım sırasında bana yardımcı olan değerli vakitlerini benim için harcayan öğretmen arkadaşlarım Zehra TAŞ, Mizamiye EKİNCİ KARAGÖZ, Merve GÜRÜN ve Beşire TUNÇ'a sonsuz teşekkürler.

Yüksek lisans eğitimlerimiz boyunca birlikte çalıştığımız, bana her zaman destek vererek pes etmememi sağlayan arkadaşım Burçin GÖRGÜLÜ ÖZTÜRK'e teşekkür ederim.

Son olarak benden hiçbir zaman desteklerini esirgemeyen, her zaman yanımda olan aileme çok teşekkür ederim.

**AÇIK UÇLU VE ÇOKTAN SEÇMELİ MADDELERDEN OLUŞAN
MADDE TAKIMLARININ GÜVENİRLİĞE ETKİSİNİN
GENELLENEBİLİRLİK KURAMI İLE İNCELENMESİ
(Yüksek Lisans Tezi)**

Serpil Kocaoğlu
GAZİ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
Eylül 2023

ÖZ

Bu çalışmada, benzer içerikli hazırlanmış açık uçlu ve çoktan seçmeli maddelerden oluşan madde takımlarının güvenilirliğe etkisini ortaya çıkarmak amaçlanmıştır. Bu amaçla, biri çoktan seçmeli diğeri açık uçlu maddelerden oluşan madde takımlarının yer aldığı iki farklı Matematik Başarı Testi hazırlanmıştır. Pilot uygulama sonucu elde edilen nihai formlar 2022-2023 Eğitim Öğretim yılında Ankara ilinde yer alan Mili Eğitim Bakanlığına bağlı bir devlet ortaokulunda 8. sınıfta öğrenim gören 128 öğrenciye uygulanmıştır. Açık uçlu maddelerin yer aldığı testte öğrencilerin 3'er maddeden oluşan 4 madde takımına verdikleri yanıtlar 3 puanlayıcı tarafından değerlendirilmiştir. Elde edilen veriler üzerinde güvenilirlik kestirimleri Genellenebilirlik Kuramı ile yürütülmüştür. Açık uçlu maddelerden oluşan madde takımlarının yer aldığı başarı testinde, madde takımı etkisinin ele alınmadığı durumda tüm öğrencilerin (b) tüm maddelere (m) verdikleri yanıtların tüm puanlayıcılar (p) tarafından puanlandığı bmxmp tümüyle çaprazlanmış desen kullanılırken; madde takımı etkisinin dahil edildiği durumda maddelerin madde takımlarında (t) yuvalandığı, bireylerin ve

puanlayıcıların ise bunlarla çaprazlandığı bx(m:t)xp yuvalanmış deseni kullanılmıştır. Elde edilen sonuçlara göre; madde takımı etkisi dahil edilmediğinde güvenilirlik katsayısı 0,81 kestirilirken, madde takımı etkisi ele alındığında güvenilirlik katsayısı 0,67 olarak kestirilmiştir. Çoktan seçmeli maddelerden oluşan madde takımlarının yer aldığı başarı testinde de benzer şekilde; madde takımı etkisinin ele alınmadığı durumda tüm öğrencilerin tüm maddelere verdikleri cevaplar bxm çaprazlanmış deseni, madde takımı etkisinin ele alındığı durumda ise tüm öğrencilerin tüm madde takımlarında yer alan tüm maddeleri cevapladığı bx(m:t) yuvalanmış deseni kullanılmıştır. Elde edilen sonuçlara göre; madde takımı etkisi dahil edilmediğinde güvenilirlik katsayısı 0,76 olarak hesaplanırken, madde takımı etkisi dahil edildiğinde güvenilirlik katsayısı 0,71 olarak kestirilmiştir. Çalışma kapsamında elde edilen verilere göre; açık uçlu maddelerin yer aldığı testte güvenilirlik katsayısı daha yüksek kestirilirken, madde takımı etkisi dahil edildiğinde güvenilirlik katsayısındaki değişim çoktan seçmeli maddelerin yer aldığı testteki değişime göre daha fazla olmuştur. Genellenebilirlik kuramı kapsamında yürütülen karar çalışmalarında madde ve madde takımı sayısındaki artışların güvenilirliğe etkisinin olumlu olduğu gözlenmekle birlikte madde takımındaki artışın güvenilirliğe katkısının daha fazla olduğu görülmüştür. Puanlayıcı sayısındaki artışın ise madde ve madde takımlarıyla kıyaslandığında güvenilirliğe katkısının daha az olduğu görülmüştür. Sonuç olarak; madde takımlarının yer aldığı testlerin güvenilirlik kestirimleri yapılırken daha yansız sonuçlar elde etmek adına madde takımı etkisinin ele alınması gerektiğine, ayrıca madde takımları içindeki madde sayısının artırılmasından ziyade madde takımı sayısının artırılmasının güvenilirliğe katkısının daha fazla olduğuna ulaşılmıştır.

Anahtar Kelimeler : Açık uçlu madde, Çoktan seçmeli madde, Madde takımı, Genellenebilirlik Kuramı, Güvenirlik
Sayfa Adedi : xvi + 109
Danışman : Doç. Dr. Melek Gülşah Şahin

**INVESTIGATION OF THE EFFECT OF TESTLETS CONSISTING
OF OPEN-ENDED AND MULTIPLE CHOICE ITEMS ON
RELIABILITY VIA GENERALIZABILITY THEORY
(M. S. Thesis)**

**Serpil Kocaoğlu
GAZI UNIVERSITY
GRADUATE SCHOOL OF EDUCATIONAL SCIENCES
September 2023**

ABSTRACT

In this study, it was aimed to reveal the effect of testlets consisting of open-ended and multiple-choice items prepared with similar content on reliability. For this purpose, two different Mathematics Achievement Tests were prepared, one consisting of multiple choice and open-ended items. The final forms obtained as a result of the pilot application were applied to 128 8th grade students in a public secondary school affiliated to the Ministry of National Education in the province of Ankara in the 2022-2023 academic year. In the test with open-ended items, the answers given by the students to the 4 item testlets consisting of 3 items were evaluated by 3 raters. Reliability estimations on the obtained data were carried out with the Generalizability Theory. In the achievement test with open-ended testlets, in the absence of the testlet effect, the $pxixr$ fully crossed design was used, in which all students' answers to (p) all items (i) were scored by all raters (r); In the case of including the testlet effect, the $px(i:h)xr$ nested design was used, in which items were nested in the testlets (h) and individuals and raters were crossed with them. According to the results obtained; The

reliability coefficient was estimated as 0.81 when the testlet effect was not included, while the reliability coefficient was estimated as 0.67 when the testlet effect was taken into account. Similarly, in the achievement test, which includes testlets consisting of multiple-choice items; In the case that the testlet effect was not addressed, the pxi crossed design was used for all the all students' answer to all the items, and in the case of the testlet effect, the px(i:h) nested design in which all the students answered all the items in all the testlets used. According to the results obtained; While the reliability coefficient was calculated as 0.76 when the testlet effect was not included, the reliability coefficient was estimated as 0.71 when the testlet effect was included. According to the data obtained within the scope of the study; While the reliability coefficient was estimated higher in the test with open-ended items, the change in the reliability coefficient was higher when the testlet effect was included, compared to the change in the test with multiple-choice items. In the decision studies carried out within the scope of generalizability theory, it was observed that the increase in the number of items and testlets had a positive effect on reliability, but it was observed that the increase in the testlet contributed more to reliability. It was observed that the increase in the number of raters contributed less to the reliability when compared to the items and testlets. In conclusion; While estimating the reliability of the tests including the testlets, it was concluded that the testlet effect should be considered in order to obtain more unbiased results, and that increasing the number of testlets, rather than increasing the number of items in the testlets, contributed more to the reliability.

Key Words : Open-ended item, Multiple choice item, Testlet, Generalizability Theory, Reliability

Page Number : xvi + 109

Supervisor :Assoc. Prof. Dr. Melek Gülşah Şahin

İÇİNDEKİLER

TELİF HAKKI VE TEZ FOTOKOPİ İZİN FORMU	i
ETİK İLKELERE UYGUNLUK BEYANI	ii
JÜRİ ONAY SAYFASI	iii
TEŞEKKÜR	v
ÖZ	vi
ABSTRACT	viii
İÇİNDEKİLER	x
TABLolar LİSTESİ	xiii
ŞEKİLLER LİSTESİ	xv
SİMGELER VE KISALTMALAR LİSTESİ	xvi
BÖLÜM I	1
GİRİŞ	1
Problem Durumu	1
Açık Uçlu Testler	3
Çoktan Seçmeli Testler	6
Madde Takımları	9
Klasik Test Kuramı	10
Genellenebilirlik Kuramı	10
<i>Genellenebilirlik Kuramının Temel Kavramları</i>	<i>11</i>
<i>Tek Yüzeyli Ölçmelerde Değişkenlik Kaynakları ve Varyans Bileşenleri</i>	<i>15</i>
<i>İki Yüzeyli Ölçmelerde Varyans Kaynakları ve Varyans Bileşenleri</i>	<i>19</i>
Araştırmanın Amacı ve Önemi	21
Problem Cümlesi	22
<i>Alt Problemler</i>	<i>22</i>

Sınırlılıklar.....	24
Sayıtlar.....	25
BÖLÜM II.....	26
İLGİLİ ARAŞTIRMALAR.....	26
Madde Takımları Üzerine Yapılan Araştırmalar.....	26
Açık Uçlu veya Çoktan Seçmeli Maddelerde Genellenebilirlik Kuramı Üzerine Yapılan Araştırmalar.....	29
BÖLÜM III	33
YÖNTEM.....	33
Araştırmanın Yöntemi.....	33
Çalışma Grubu.....	33
Veri Toplama Aracı.....	35
Uygulama Süreci	40
Verilerin Analizi.....	40
BÖLÜM IV	46
BULGULAR VE YORUMLAR.....	46
Birinci Alt Probleme İlişkin Bulgular ve Yorumlar	46
İkinci Alt Probleme İlişkin Bulgular ve Yorumlar.....	54
Üçüncü Alt Probleme İlişkin Bulgular ve Yorumlar.....	60
BÖLÜM V	62
SONUÇ, TARTIŞMA VE ÖNERİLER	62
Birinci Alt Probleme İlişkin Sonuç ve Tartışma	62
İkinci Alt Probleme İlişkin Sonuç ve Tartışma.....	64
Üçüncü Alt Probleme İlişkin Sonuç ve Tartışma.....	66
Öneriler.....	67
KAYNAKLAR.....	69
EKLER	75
EK 1. Pilot Uygulama Sonucu Elde Edilen Çoktan Seçmeli Maddelerden Oluşan Madde Takımlarının Yer Aldığı Başarı Testine Ait Tap Programı Analiz Çıktıları..	76
EK 2. Çoktan Seçmeli Maddelerden Oluşan Madde Takımlarının Yer Aldığı Matematik Başarı Testinin Nihai Formu	78
EK 3. Açık Uçlu Maddelerden Oluşan Madde Takımlarının Yer Aldığı Matematik Başarı Testinin Nihai Formu	82
EK 4. Açık Uçlu Maddelerin Yer Aldığı Başarı Testine Ait Dereceli Puanlama Anahtarı	88

EK 5. Nihai Uygulama Sonucu Elde Edilen Çoktan Seçmeli Maddelerden Oluşan Madde Takımlarının Yer Aldığı Başarı Testine Ait TAP Programı Analiz Çıktıları	96
EK 6. EduG Programı Analiz Çıktıları	98
Çoktan Seçmeli Maddelerden Oluşan Başarı Testinde Madde Takımı Etkisinin Ele Alınmadığı $b \times m$ Desenine Ait Analiz Çıktıları	98
Çoktan Seçmeli Maddelerden Oluşan Başarı Testinde Madde Takımı Etkisinin Dikkate Alındığı $b \times (m:t)$ Desenine Ait Analiz Çıktıları	99
Açık Uçlu Maddelerden Oluşan Başarı Testinde Madde Takımı Etkisinin Ele Alınmadığı $b \times m \times p$ Desenine Ait Analiz Çıktıları	101
Açık Uçlu Maddelerden Oluşan Başarı Testinde Madde Takımı Etkisinin Dikkate Alındığı $b \times (m:t) \times p$ Desenine Ait Analiz Çıktıları.....	103
EK 7. Etik Kurul İzni	106
EK 8. MEB Araştırma İzni	108

TABLÖLAR LİSTESİ

Tablo 1.1 Çoktan Seçmeli Madde Yazım Kuralları.....	9
Tablo 1.2 Tek Yüzeyli Ölçmelerde Değişkenlik Kaynakları.....	17
Tablo 3.1 Maddelerin İçerik Yönünden Benzerliğine Dair Uyum Yüzdeleri.....	37
Tablo 3.2 Pilot Uygulamadaki Testlerin Madde İstatistikleri.....	38
Tablo 3.3 Nihai Uygulamadaki Testlerin Madde İstatistikleri.....	41
Tablo 4.1 Açık Uçlu Maddelerden Oluşan Başarı Testinde Madde Takımı Etkisinin Ele Alınmadığı $b \times m \times p$ Desenine Ait G Çalışması Sonuçları.....	47
Tablo 4.2 Açık Uçlu Maddelerden Oluşan Başarı Testinde Madde Takımı Etkisinin Ele Alınmadığı $b \times m \times p$ Desenine Ait K Çalışması Sonuçları	49
Tablo 4.3 Açık Uçlu Maddelerden Oluşan Başarı Testinde Madde Takımı Etkisinin Dikkate Alındığı $b \times (m:t) \times p$ Desenine Ait G Çalışması Sonuçları	50
Tablo 4.4 Açık Uçlu Maddelerden Oluşan Başarı Testinde Madde Takımı Etkisinin Dikkate Alındığı $b \times (m:t) \times p$ Desenine Ait K Çalışması Sonuçları	53
Tablo 4.5 Çoktan Seçmeli Maddelerden Oluşan Başarı Testinde Madde Takımı Etkisinin Ele Alınmadığı $b \times m$ Desenine Ait G Çalışması Sonuçları	54
Tablo 4.6 Çoktan Seçmeli Maddelerden Oluşan Başarı Testinde Madde Takımı Etkisinin Ele Alınmadığı $b \times m$ Desenine Ait K Çalışması Sonuçları	56
Tablo 4.7 Çoktan Seçmeli Maddelerden Oluşan Başarı Testinde Madde Takımı Etkisinin Dikkate Alındığı $b \times (m:t)$ Desenine Ait G Çalışması Sonuçları	57
Tablo 4.8 Çoktan Seçmeli Maddelerden Oluşan Başarı Testinde Madde Takımı Etkisinin Dikkate Alındığı $b \times (m:t)$ Desenine Ait K Çalışması Sonuçları	59

Tablo 4.9 <i>Açık Uçlu ve Çoktan Seçmeli Maddelerden Oluşan Başarı Testlerinde Madde Takımı Etkisine Dayalı Elde Edilen Güvenirlik Katsayıları</i>	60
--	----



ŞEKİLLER LİSTESİ

<i>Şekil 1.1.</i> Puanlayıcı etkileri.....	5
<i>Şekil 1.2.</i> Bir yüzeyli çaprazlanmış bxm deseni için değişkenlik kaynakları ve varyans bileşenleri.....	18
<i>Şekil 1.3.</i> Bir yüzeyli yuvalanmış m:b deseni için değişkenlik kaynakları ve varyans bileşenleri.....	19
<i>Şekil 1.4.</i> İki yüzeyli çaprazlanmış bxgxp deseni için varyans kaynakları ve varyans bileşenleri.....	20
<i>Şekil 1.5.</i> İki yüzeyli kısmi yuvalanmış bx(m:p) deseni için varyans kaynakları ve varyans bileşenleri.....	21
<i>Şekil 3.1.</i> İki yüzeyli çaprazlanmış bxm _x p deseni için varyans kaynakları ve varyans bileşenleri.....	42
<i>Şekil 3.2.</i> Üç yüzeyli bx(m:t) _x p deseni için varyans kaynakları ve varyans bileşenleri.....	43
<i>Şekil 3.3.</i> İki yüzeyli yuvalanmış bx(m:t) deseni için varyans kaynakları ve varyans bileşenleri... ..	44

SİMGELER VE KISALTMALAR LİSTESİ

KTK: Klasik Test Kuramı

G Kuramı: Genellenebilirlik Kuramı

G Çalışması: Genellenebilirlik Çalışması

K Çalışması: Karar Çalışması

G Katsayısı: Genellenebilirlik Katsayısı

Phi Katsayısı: Güvenirlilik Katsayısı

BÖLÜM I

GİRİŞ

Bu bölümde araştırmanın; problem durumu, amacı ve önemi, sınırlılıkları ile sayıltılarına yer verilmiştir.

Problem Durumu

Eğitimde birden çok amaca dönük olarak kararlar alınmaktadır. Öğrenci başarısının belirlenmesi ve izlenmesi, öğrencilerin programa seçimi, program değerlendirme, araştırma yapma gibi amaçlara yönelik pek çok değişken değerlendirilebilir (Turgut, 1992). Değerlendirme, ölçme sonuçlarını ölçüt ya da ölçütlerle karşılaştırarak bir değer yargısına varma sürecidir (Doğan, 2019; Tekin, 2009; Turgut & Baykul, 2010). Bilindiği gibi eğitimde değerlendirilen değişkenlerin çoğu öğrenci nitelikleri ile ilgilidir. Bu nitelikler; öğrencilerin başarı, hazır bulunuşluk, ilgi, tutum, öz yeterlik, motivasyon, kaygı, deney yapma, resim yapma, koşma, materyal tasarlama vb. gibi bilişsel, duyuşsal ve psikomotor özelliklerdir. Belirtilen özellikler bireyler arasında farklılık gösterdiğinden ölçme işlemi bu farklılıkları belirlemeye dönük olarak gerçekleştirilir. Ölçme işlemi; bir betimleme işi olup, bir niceliğin belli bir özelliğe sahip olma düzeyinin gözlenip, gözlem sonuçlarının sayı veya sembollerle

ifade edilmesidir (Tekin, 2009). Eğitim alanında yapılan ölçme işlemleri genellikle ölçme araçları ile gerçekleştirilmektedir.

Eğitimde ölçme araçları ile gerçekleştirilen ölçümlerin amacına dönük ve hatasız olabilmesi için ölçme aracında birtakım özellikler göz önüne alınmalıdır. Bir ölçme aracında olması gereken üç temel nitelik vardır: Güvenirlik, geçerlik ve kullanılabilirlik. Güvenirlik, ölçme aracının tesadüfî hatalardan arınlık derecesini ifade eder. Ölçümlerde mutlaka bir miktar hata karışacağı göz önüne alınırsa güvenirlik, bu hatalara rağmen ne kadar doğru bir ölçüm yapıldığının göstergesi olacaktır. Ölçme aracında olması gereken en önemli nitelik ise geçerliktir. Geçerlik; bir ölçme aracının ölçmek istediği özelliği, farklı değişkenler karıştırmadan ölçebilme derecesi olarak tanımlanmaktadır (Başokçu, 2019; Kan, 2009; Tekin, 2009; Turgut & Baykul, 2010). Geçerlik ve güvenirlik bir ölçme aracının olmazsa olmazlarıdır. Tek başına ne sadece güvenirlik ne de geçerlik yeterlidir. Dolayısıyla bu iki nitelik mutlaka incelenmelidir. Bir ölçme aracında olması gereken diğer bir nitelik ise kullanılabilirliktir. Tekin (2009) bir testin kullanılabilirliğini; testin geliştirilmesinin, çoğaltılmasının, uygulanmasının ve puanlanmasının kolay ve ekonomik olması olarak ifade etmektedir. Bu sebeple bir ölçme aracı hazırlanırken kullanılabilir bir ölçme aracı hazırlanması için dikkate alınması gerekenler şu şekildedir: Ekonomiklik, hazırlama ve uygulama süresi, hazırlayıcı ve uygulayıcının nitelikleri, cevaplayıcının nitelikleri, uygulama ve puanlama kolaylıkları ile puanları yorumlama kolaylıkları (Turgut & Baykul, 2010).

Eğitimde kalitenin de önemli göstergelerinden birisi olan ölçme araçlarından elde edilen bilgilerin geçerli ve güvenilir olarak yorumlanabilmesi eğitimde alınacak kararlar açısından oldukça önemlidir. Bu nedenle öğrenci başarısının belirlenmesinde önemli bir rolü olan başarı testlerinin hazırlanmasında bir takım uyulması gereken hususlar söz konusudur. Öncelikle şunu bilmek gerekir ki test hazırlama bir süreç işidir ve bu sürecin her aşaması titizlikle planlanmalıdır. Bir test hazırlanırken; testin amacı, testin kapsamı, test maddelerinin hazırlanması, uygulanması ve değerlendirilmesi aşamaları göz önünde bulundurulmalıdır (Berberoğlu, 2006; Özçelik, 2013).

Başarı testleri genel olarak standart testler ve öğretmen yapımı testler olarak ikiye ayrılmaktadır (Tekin, 2009). Standart testler daha çok ulusal ve uluslararası sınavlarda kullanılan, uzmanlar tarafından hazırlanan testler iken, öğretmen yapımı testler adından da anlaşılacağı üzere öğretmenin kendisinin hazırladığı, bir dersin ya da konunun amaçları için geliştirilmiş testlerdir. Öğretmen yapımı testlerde en çok tercih edilen madde türlerinden birinin çoktan seçmeli maddeler (Çakan, 2004) olması ile birlikte literatürde açık uçlu, kısa cevaplı, eşleştirmeli maddeler ve doğru-yanlış maddeleri de yer almaktadır. Bu araştırmada açık uçlu ve çoktan seçmeli madde takımlarından oluşan testlerin güvenilirliklerinin incelenmesi amaçlandığından aşağıda açık uçlu testler, çoktan seçmeli testler ve madde takımlarına değinilmiştir.

Açık Uçlu Testler

Açık uçlu maddelerden oluşan testlerde öğrencilere bir ya da birkaç madde verilir ve öğrencilerin bu madde/maddelere yazılı olarak cevap vermesi beklenir. Öğrencilerden maddelere bazen kısa ve kesinlikle belirlenmiş bir cevap vermesi beklenirken bazen de cevabın niteliği ve uzunluğu açısından öğrenci serbest bırakılabilir. Bu bağlamda ele alındığında açık uçlu maddeler yanıtı sınırlandırılmış ve serbest bırakılmış olmak üzere iki başlıkta toplanır (Berberoğlu, 2006).

Genel olarak açık uçlu madde hazırlanmak istendiğinde dikkat edilmesi gereken bazı noktalar bulunur. Aşağıda açık uçlu madde yazarken dikkat edilmesi gereken hususlara kısaca değinilmiştir (Doğan, 2009a; Tekin, 2009; Turgut & Baykul, 2010).

- Öncelikle her testte olduğu gibi burada da test planı dikkatli bir şekilde yapılmalıdır. Testin amacına uygun olacak şekilde kapsam geçerliğini de düşünerek maddelerin hazırlanması tercih edilmelidir.
- Maddeler açık, anlaşılır ve net olmalıdır. Her öğrenci maddeden farklı bir anlam çıkarmamalı ve maddeler farklı yönlerle çekilebilir olmamalıdır.

- Hem kapsam geçerliğini sağlamak hem de güvenilirliği artırmak adına uzun cevaplı az sayıda madde kullanmak yerine daha kısa cevaplı çok sayıda madde kullanılmalıdır.
- Maddeler birbirinden bağımsız olmalı, bir maddenin cevabı diğer bir maddenin içinde olmamalı ya da onun çözümünü etkilememelidir.
- Sınavda seçmeli madde bulundurulmamalıdır. Bazen öğretmenler sınavda seçmeli madde sorup öğrencilerden istediğini cevaplamasını beklerler. Bu durumun olumsuz sonuçları vardır. Öğrenciler seçim yaptığında bazı davranışları ölçen maddeleri cevaplamamış olacaktırlar. Aynı zamanda seçilen maddelerin puanlamasında kıyaslama hatası olabileceğinden güvenilirliği olumsuz yönde etkileyecektir.
- Açık uçlu maddeler yazılırken yazım kurallarına dikkat edilmelidir.
- Öğrenciler için bir yönerge eklenmelidir.
- Açık uçlu maddelerin en büyük dezavantajı puanlama kısmıdır. Bunu minimuma indirmek için çok iyi bir puanlama anahtarı hazırlanmalıdır. Yanıtı sınırlı olan açık uçlu maddelerde ölçütler net bir şekilde belirlenmeli, her bir maddeye her bir davranışa kaç puan verileceği önceden belirlenmelidir. Serbest cevaplı maddelerde ise öğrenciler daha özgür olacaklarından sınıflama yöntemi ile puanlama yapmak daha uygun olacaktır. Yanıtlar; tam doğru, kısmi doğru, boş ve yanlış yanıtlar şeklinde derecelendirilebilir. Ayrıca bu tarz testlerde birden fazla puanlayıcı kullanılması da tercih edilmektedir.

Açık uçlu maddelerin öğretmenler arasında kullanımı oldukça yaygındır. Özellikle üst düzey bilişsel becerileri ölçmek için uygun bir araç olduğundan sıklıkla tercih edilmektedir. Diğer yandan ele alındığında ise açık uçlu maddelerin puanlanmasında genel olarak puanlamaya hata karışacağı görüşü hakimdir. Öğrencinin açık uçlu maddelere verdiği yanıtlardan elde ettiği başarı puanını etkileyen puanlayıcıdan kaynaklı bu faktörler puanlayıcı etkileri olarak tanımlanmaktadır. Performans değerlendirmeye karışan bu puanlayıcı etkileri Şekil 1.1.'de gösterilmiştir (Çetin & İlhan, 2017; Eckes, 2017; Saal, Downey & Lahey, 1980; Şata, 2019).

Puanlayıcı katılığı ve cömertliği: Puanlayıcının diğer puanlayıcılardan ya da puanlama ölçütlerinden daha düşük/yüksek puanlama yapma eğilimidir.

Merkeze yönelme etkisi: Puanlayıcının puanlama ölçeğindeki orta kategoriye diğerlerine göre daha fazla eğilim göstermesidir.

Halo etkisi: Puanlayıcının bireyin farklı özelliklerine aynı puanları verme eğilimidir.

Ranj sınırlaması: Puanlayıcının bireylerin performanslarını birbirinden ayıramama durumudur.

Yanlılık ve tutarsızlık: Puanlayıcının öğrencilerden bazılarına daha katı ve cömert davranma eğilimidir.

Şekil 1.1. Puanlayıcı etkileri.

Öğrencilerin performanslarının değerlendirilmesinde Şekil 1.1’de de görüldüğü gibi birçok puanlayıcı etkisi ile karşı karşıya kalınmaktadır. Böyle bir durumda puanlayıcı güvenilirliği ön plana çıkmaktadır. Bundan dolayı puanlamanın objektif olması önem arz ettiği için kurallara uygun olarak hazırlanmış dereceli puanlama anahtarı kullanılmalıdır. Dereceli puanlama anahtarı bir diğer adıyla rubrik; genellikle belirli bir çalışma için belirlenen özelliğin mükemmel olduğu durumdan hiç gözlenmediği duruma doğru değişen düzeyleri gösteren bir doküman olarak tanımlanmaktadır (Andrade, 2000; Goodrich, 1996). Dereceli puanlama anahtarı ile maddelerin yanıtları tam doğru, kısmi doğru ve boş ya da yanlış yanıtlar başlığı altında gruplandırılabilir. Bu sayede özellikle açık uçlu maddelerin puanlanması hem daha objektif olacak hem de daha az zaman alacaktır.

Çoktan Seçmeli Testler

Çoktan seçmeli maddelerden oluşan testler en sık kullanılan öğretmen yapımı testlerdir. Genel olarak çoktan seçmeli madde, madde kökü ve seçeneklerden oluşmaktadır. Seçeneklerden biri doğru yanıt iken diğer seçenekler ise çeldiriciler olarak adlandırılmaktadır (Doğan, 2009b; Tekin, 2009).

Çoktan seçmeli maddelerde bireylerin açık uçlu maddelerde olduğu gibi cevabı kendilerinin yazması beklenmemekte, kendilerine sunulan seçeneklerden doğru olduğunu düşündükleri seçeneği işaretlemeleri beklenmektedir. Bu sebeple cevaplama süresi daha kısadır ve açık uçlu maddelere kıyasla daha fazla sayıda madde testte yer alabilir. Böylece çoktan seçmeli testlerde kapsam geçerliğinin daha iyi sağlandığı görülmektedir. Açık uçlu testlerin aksine puanlama objektif bir şekilde kolaylıkla yapılabilir. Ancak açık uçlu maddelere kıyasla sınırlılıklarından birisi ise çoktan seçmeli test maddesi hazırlamanın daha zor olduğu belirtilebilir. Çoktan seçmeli madde yazımı özel bilgi, beceri ve donanım gerektirmektedir. Doğru yanıtın kesin olarak belirlenmesi, çeldiricilerin ise çok iyi hazırlanmış olması beklenir. Açık uçlu testlerde üst düzey bilişsel becerileri ölçmek mümkün iken, çoktan seçmeli testlerde daha çok alt düzey becerilerin ölçüldüğü düşünülmektedir. Ancak çok iyi hazırlanmış bir çoktan seçmeli madde ile uygulama, analiz etme gibi üst düzey bilişsel beceriler de ölçülebilir.

Çoktan seçmeli maddeler farklı formatlarda hazırlanabilir. En çok kullanılan türleri aşağıda belirtilmiştir (Doğan, 2009b; Doğan, 2019; Haladyna, 2004; Tekin, 2009; Turgut & Baykul, 2010).

1. Doğru yanıtı göre;
 - a. Tek bir doğru cevap isteyen maddeler: Madde kökünde sorulan sorunun tek bir doğru cevabı vardır, diğer seçenekler kesinlikle yanlıştır.
 - b. En doğru cevap isteyen maddeler: Verilen seçeneklerin hepsinde doğruluk payı vardır. Ancak seçeneklerden birinde en doğru cevap bulunurken, diğer seçeneklerde kısmen doğru cevaplar yer almaktadır.

- c. Bileşik yanıt isteyen maddeler: Bu maddelerde genellikle kökteki sorunun yanıtı birden fazladır. Bu yüzden çoğunlukla madde kökünde öncüller kullanılmaktadır.
- d. Doğru yanıt gizlenen maddeler: Madde kökünde sorulan sorunun cevabının açık bir şekilde ifade edilmesi, bireylerin yanıtı görüp hatırlamasını kolaylaştırabilir. Bu nedenle seçeneklerde yanıtlar gizlenmiş bir şekilde verilmektedir.
- e. Doğru yanıt birden fazla olan maddeler: Madde kökünün tek bir doğru cevabı yoktur. Maddenin seçenekleri arasında iki veya daha fazla doğru cevap bulunur.

2. Madde köküne göre;

- a. Madde kökü soru kipinde olan maddeler
 - i. Olumlu soru biçimindeki maddeler: Madde kökünde olumlu soru kipi bulunmaktadır.
 - ii. Olumsuz soru biçimindeki maddeler: Madde kökü olumsuz kelimelerle bitmektedir.
- b. Kökü eksik cümle tipinde olan maddeler: Madde kökü verilen seçeneklerden her biri ile tamamlanabilecek şekilde eksik cümle tipinde bırakılmaktadır.

3. Maddelerin gruplanışına göre;

- a. Ortak köklü maddeler: Bir paragraf, tablo, grafik, şekil vb. gibi materyallerle sorulan iki ya da daha fazla maddeler ortak köklü madde olarak ifade edilir.
- b. Ortak seçenekli maddeler: Bazı maddelerde aynı seçenekler kullanılabilir. Bu durumda öncelikle ortak seçenekler üstte verilip daha sonra maddeler sıralanabilir.

Çoktan seçmeli maddenin formatı belirlendikten sonra, madde yazımında dikkat edilmesi gereken birtakım kurallar bulunmaktadır (Doğan, 2019; Haladyna, 2004; Tekin, 2009; Turgut & Baykul, 2010). Çoktan seçmeli maddelerin yazılmasında dikkat edilmesi gereken

kurallar Haladyna (2004) de yer alan genel madde yazım kurallarından çoktan seçmeli maddeler ile ilgili olanlar uyarlanarak Tablo 1.1’de özetlenmiştir.

Tablo 1.1

Madde Yazım Kuralları

İçerik
1. Her madde tek bir bilişsel süreci yansıtmalıdır. 2. Her madde önemli bir davranışı yoklamalı, önemsiz ayrıntılardan kaçınılmalıdır. 3. Bilgi ve becerilerin ölçülmesinde kullanılacak olan her madde öğrenci için yeni olmalıdır. 4. Her test maddesi diğer maddelerden bağımsız olarak yanıtlanabilmelidir. 5. İçeriği çok dar ya da çok kapsamlı olmamalıdır. 6. Maddelerin içeriği öznel olmamalıdır. 7. Test maddelerinde hileli, aldatıcı öğelerden kaçınılmalıdır.
Stil ve biçim
8. Seçenekleri yan yana yazmak yerine alt alta yazmak tercih edilmelidir. 9. Maddenin açık ve net olmasına dikkat edilmelidir. 10. Madde yazımında dil bilgisi, noktalama işaretleri ile yazım ve imla kurallarına dikkat edilmelidir. 11. Her madde yalın ve anlaşılabilir bir dille yazılmalıdır. 12. Madde yazımında okuma süresi göz önünde bulundurulmalı, gereksiz sözcük kullanımından kaçınılmalıdır. 13. Testteki her öğenin düzenine dikkat edilmelidir.
Madde kökü
14. Yönerge olabildiğince açık ve net olmalıdır. 15. Madde kökü olabildiğince kısa tutulmalı, gereksiz ifade ve açıklamalardan kaçınılmalıdır. 16. Madde kökü ne sorulduğunu seçeneklere bakılmadan açıkça ortaya koymalıdır. 17. Madde kökünde olumsuz sözcük kullanımından kaçınılmalıdır.
Seçenekler
18. Seçenekler mümkün olduğunca kapsam bakımından birbirine benzer olmalıdır. 19. Doğru yanıtlar seçeneklere rastgele dağıtılmalıdır. Aynı seçeneğe art arda çok sayıda doğru yanıt getirilmemelidir. 20. Seçeneklerin yazımında mantıksal veya sayısal bir sıra takip edilmelidir. 21. Seçenekler birbirinden bağımsız olmalıdır. 22. Seçenekler içerik ve dil bilgisi bakımından benzer olmalıdır. 23. Seçeneklerin uzunluğu yaklaşık olarak aynı olmalıdır. 24. Seçeneklerde “yukarıdakilerin hiçbirisi”, “yukarıdakilerin hepsi” gibi ifadelerden kaçınılmalıdır. 25. Seçeneklerde olumsuz ifade kullanmaktan kaçınılmalıdır. 26. Doğru cevabı ipucu veren seçenek yazılmamalıdır. 27. Çeldiricilerin çeldiricilik düzeylerinin benzer olmasına dikkat edilmelidir. 28. Çeldirici yazarken öğrencilerin yapabileceği tipik hatalar göz önünde bulundurulmalıdır.

“Developing and validating multiple-choice test items” Haladyna, T. M. 2004, Taylor & Francis Group, s.99 kaynağından uyarlanmıştır.

Madde Takımları

Madde takımı kavramını ilk olarak Kiely ve Wainer (1987) ortaya atmıştır. Çalışmalarında bilgisayar ortamında uyarlanmış testlerde (Computerized Adaptive Testing - CAT) içeriği dengelemek, bağlam etkisi ve madde sırasının önemini ortadan kaldırmak için madde takımlarından yararlanmışlardır.

Tekin'e (2009) göre madde takımı kavramı ortak bir materyale bağlı olarak cevaplanması gereken maddeler olarak tanımlanmaktadır. Bu materyal bir grafik, bir tablo, bir şekil, bir harita gibi farklı seçenekler olabilmektedir. Ortak materyalle ilgili en az iki ya da daha çok madde sorulabilmektedir. Yaman'a (2016) göre ise; resim, hikâye, şiir, grafik, fotoğraf, deney düzeneği gibi ortak bir içerikle ilgili 2 ile 12 maddeden oluşan madde takımı oluşturulabilir.

Madde takımları özellikle son yıllarda hem ulusal hem de uluslararası sınavlarda sıklıkla kullanılmaktadır. Ortak bir materyale bağlı olarak madde hazırlamak zaman ve enerji tasarrufu sağladığından madde yazarları tarafından tercih edilmektedir (Wainer, Bradlow & Du, 2000). Ayrıca ortak bir köke bağlı olarak ardışık maddelerin takip edilmesinin daha başarılı olduğu görülmüştür (Lee, Brennan & Frisbie, 2000). Üst düzey bilişsel becerileri de ölçmeyi mümkün kılması madde takımlarının popüler hale gelmesinde etkilidir. Madde hazırlanmasında nasıl birtakım kurallara dikkat edilmesi gerekiyorsa, madde takımı hazırlanırken de dikkat edilmesi gereken hususlara aşağıda kısaca değinilmiştir (Tekin, 2009).

- Seçilen materyal ölçülmek istenen davranışla uyumlu, açık, net ve öğrenciler için yeni ve anlaşılır olmalıdır.
- Maddeler sadece materyale bağlı olarak doğru cevaplandırılabilmelidir. Aksi halde materyal kullanımı bir anlam ifade etmeyecektir.
- Materyale ait madde sayısı dikkatli bir şekilde tercih edilmelidir.

Madde takımlarının hazırlanmasından sonraki aşamada testlerin uygulanması ve analiz edilmesi gerekmektedir. Hazırlanan maddelerin amaca hizmet edip etmediği, ölçümlerin ne derece tutarlı olduğu ise madde analizleri ile sağlanmaktadır. Bir teste ait madde istatistikleri farklı test kuramları ile ele alınabilmektedir. Bu araştırmada pilot uygulama ve nihai uygulamadan elde edilen madde istatistikleri klasik test kuramına dayalı hesaplanmıştır. Araştırma problemine dayalı güvenilirlik analizleri genellenebilirlik kuramı çerçevesinde gerçekleştirilmiştir. Aşağıda klasik test kuramı ve genellenebilirlik kuramı hakkında bilgi sunulmuştur.

Klasik Test Kuramı

Ölçmede en sık kullanılan ve en eski kuramlardan biri olan klasik test kuramı ilk olarak 1900’lü yılların başında Charles Spearman tarafından ortaya atılmıştır. Sperman’ın ortaya koyduğu bu kuramda gözlenen puan (X); bireyin gerçek puanı (T) ile tesadüfi hatanın (E) toplamından oluşmaktadır. Bu durum “ $X=T+E$ ” eşitliği ile ifade edilmiştir (Crocker & Algina, 2008). Tekrarlı ölçümler sonucu elde edilen puanların tutarlılığı olarak kabul edilen güvenilirlik, klasik test kuramında hata kaynağına göre farklılık gösterecektir. Fakat hata kaynakları burada tek yapı altında toplandığı için güvenilirlik katsayısı için farklı hata kestirimleri yapmak söz konusu olmayacaktır. Bu durum da klasik test kuramının sınırlılığı olarak belirtilebilir.

Genellenebilirlik Kuramı

Shavelson ve Webb’e (1991) göre genellenebilirlik (G) kuramı, davranışsal ölçümlerin güvenilirliği hakkında fikir veren istatistiksel bir kuramdır. Hem klasik test kuramının hem de varyans analizinin bir uzantısı olarak, çoklu hata kaynaklarının ele alınabildiği matematiksel bir modeldir.

G kuramının üstünlüğü, tek bir analizle farklı hata kaynaklarının aynı anda kestirilebilmesidir. Diğer bir deyişle klasik test kuramının aksine, farklı hata kaynaklarının hem ayrı ayrı hem de tek bir etkileşimle ortaya çıkardıkları sonuçları göz önüne almaktadır. G kuramı ayrıca, puanların güvenilirliğini farklı yorumlamalar için kestirebilmeyi de sağlamaktadır. Klasik test kuramında bireylerin sadece birbirlerine göre durumlarının karşılaştırıldığı göreceli (bağıl) değerlendirme yapılırken; G kuramında bireylerin birbirinden bağımsız olarak sadece performanslarının değerlendirildiği mutlak değerlendirmeler yapmak da mümkündür. Yani; klasik test kuramı araştırmacılara yalnızca göreceli karar verebilecekleri bilgi sağlarken, G kuramı hem göreceli hem de mutlak kararlar için yeterli bilgiyi aynı anda sunar (Brennan, 2001; Güler, Kaya Uyanık & Taşdelen Teker, 2012; Shavelson & Webb, 1991). G kuramının daha iyi anlaşılabilmesi için bazı temel kavramlara aşağıda yer verilmiştir.

Genellenebilirlik Kuramının Temel Kavramları

- 1. Yüzey, Koşul ve Evren:** Klasik test kuramında güvenilirlik hesaplamalarında tek bir hata kaynağı ele alınırken G kuramında hata kaynaklarının birden fazla olduğu belirtilmişti. G kuramında ele alınan madde, puanlayıcı, form, zaman gibi bu hata kaynakları “yüzey (facet)” olarak adlandırılır. Bir başka deyişle G kuramında değişkenlik kaynaklarına yüzey adı verilmektedir. Her bir yüzey de kendi içinde düzeylerden oluşmaktadır. Değişkenlik kaynaklarını oluşturan bu düzeyler de “koşul” olarak adlandırılmaktadır. Örneğin puanlayıcı bir yüzey olarak alındığında; birinci, ikinci ve üçüncü puanlayıcının her biri bu yüzeyin koşulunu oluşturur. G kuramındaki yüzey ve koşul kavramları ile tek yönlü varyans analizinde (ANOVA) faktör ve düzey olarak karşılaşılmaktadır. (Atılğan, 2019; Brennan, 2001; Crocker & Algina, 2008; Güler, Kaya Uyanık & Taşdelen Teker, 2012; Shavelson & Webb, 1991).

Ölçme desenindeki değişkenlik kaynaklarının koşullarına karşılık gelen ve bütün bu koşullardan elde edilebilecek ölçme sonuçları evrenine “kabul edilebilir gözlemler evreni” denir. Çalışmada bir yüzeyin koşullarından genelleme yapılmak istenen evrenler de “genelleme evrenini” oluşturmaktadır (Atılğan, 2019; Brennan, 2001; Crocker & Algina, 2008; Shavelson & Webb, 1991). Örneğin 12 açık uçlu maddeye verilen cevapların 3 farklı puanlayıcı tarafından puanlandığı bir durumda olası tüm gözlemler evreni madde ve puanlayıcı yüzeylerinden oluşmaktadır. Başka bir ifadeyle kabul edilebilir gözlemler evreni olası herhangi bir madde ile olası herhangi bir puanlayıcı içerebilir. Olası üç puanlayıcı ve 12 maddeden oluşan tüm durumlar ise genelleme evrenini oluşturmaktadır.

2. Ölçme Objesi: Birçok ölçme durumunda bireyler yani sınava giren adaylar ölçme objesi olarak belirlenmektedir. Bir başka ifadeyle ölçme objesi ölçmenin hedefleri durumundadır. Bireyler arası farklılıklar insanın doğası gereği var olan bir durum olduğundan klasik test kuramında hata kaynağı olarak ele alınmadığı gibi G kuramında da değişkenlik kaynağı olarak düşünülmez. Ölçme objesine ait varyans bileşeni bize bireyler arasındaki farklılığı gösterir. Fakat her ölçme durumunda ölçme objesinin birey olma zorunluluğu yoktur. Başka bir ölçmede madde gibi farklı ölçme objeleri de olabilir (Brennan, 2001; Crocker & Algina, 2008; Güler, Kaya Uyanık & Taşdelen Teker, 2012; Shavelson & Webb, 1991).

3. Genellenebilirlik (G) ve Karar (K) Çalışmaları: G kuramında güvenilirlik kestirimleri genellenebilirlik (G) ve karar (K) çalışmaları ile yürütülmektedir. G çalışmasında amaç, olası bütün değişkenlik kaynaklarını analiz ederek hata kaynaklarını belirlemek ve etkilerini ortaya çıkararak kabul edilebilir gözlemler evrenini tanımlamaktır. Bu bağlamda G çalışması; ölçmedeki varyans kaynakları hakkında olabildiğince bilgi sağlayarak ölçme hakkında kararlar verme, değişkenlik kaynaklarından gelen hataları belirleyerek bunların azaltılması için çalışmalar yapma konularında bilgi sağlar (Atılğan, 2019; Brennan, 2001; Crocker & Algina, 2008). G

çalışmasının bir diğer amacı da K çalışmasına veri sağlamaktır. K çalışmasının temel amacı, G çalışmasından elde edilen verileri kullanarak belirli bir amaç için karar vermektir. Bir başka deyişle K çalışmalarında yapılan ölçmenin en uygun desenine ulaşılmasıdır. Bir sınava yeni madde eklenmesinin ya da sınavdan madde çıkartılmasının, puanlayıcı sayısının artırılıp azaltılmasının güvenilirliği nasıl etkilendiğine K çalışması ile bakılır. Bu sayede en uygun ölçme deseni elde edilmesi sağlanır (Atılğan, 2019; Güler, Kaya Uyanık & Taşdelen Teker, 2012).

- 4. Çaprazlanmış ve Yuvalanmış Desen:** G kuramında desenin çaprazlanmış (crossed) ya da yuvalanmış (nested) olarak belirlenmesi analizlerde farklılıklara yol açacaktır. Çaprazlanmış desenlerde, bir yüzeyin bütün koşulları diğer bir yüzeyin bütün koşullarıyla örtüşmektedir ve bu durum yüzeyler arasına “x” işareti konularak gösterilir (Atılğan, 2019; Brennan, 2001; Shavelson & Webb, 1991). Örneğin açık uçlu 12 maddeden (m) oluşan bir başarı testini 128 öğrencinin (b) her birinin yanıtladığı ve bu yanıtların her birinin 3 puanlayıcı (p) tarafından puanlandığında desen $b \times m \times p$ tümüyle çaprazlanmış desen olarak tanımlanabilir.

Bir yüzeyin koşulları diğer bir yüzeyin bazı koşulları ile örtüşüyorsa bu durumda desen yuvalanmış olarak adlandırılır ve “:” işareti konularak gösterilir (Atılğan, 2019; Brennan, 2001; Shavelson & Webb, 1991). Örneğin her madde takımında (t) 3 maddenin (m) yer aldığı ve bu maddelere verilen cevapların 3 puanlayıcı (p) tarafından puanlanması durumunda karşımıza çıkan desen yuvalanmış desen olarak tanımlanabilir.

- 5. Bağıl ve Mutlak Değerlendirme:** G çalışmaları araştırma yapacak olan kişi bağıl (görelî) veya mutlak modeller kullanarak bağıl veya mutlak değerlendirmelerde bulunabilir. Bağıl değerlendirmelerde genel olarak bireylerin ölçme sonuçlarının ölçmenin uygulandığı grup içinde sıralanması söz konusudur. Başka bir ifadeyle bir öğrencinin başarı sırası diğer arkadaşlarının da puanına bağıl olarak değişecektir. Mutlak değerlendirmede ise, bireyin almış olduğu puan daha önceden belirlenmiş bir

ölçütü kıyaslanmaktadır. Örneğin bir sınavda 100 tam puan üzerinden 70 puan alanlar başarılı sayılacaksa burada mutlak değerlendirme yapılmış olacaktır. Önceden belirlenmiş bir kriter söz konusudur. Sonuç olarak; bağıl değerlendirmede grup puanlarından yola çıkarak bir ölçüt belirlenirken, mutlak değerlendirmede gruptan bağımsız önceden belirlenmiş bir ölçüt bulunmaktadır (Atılğan, 2019).

6. Tesadüfi ve Sabit Yüzeyler: Tesadüfi ve sabit yüzey, yüzeylerin koşuluna uygun olarak seçilen örneklem ile ilgilidir. Örneklem büyüklüğü evren büyüklüğünden çok daha küçük, örneklem evrenden rastgele alınmış veya evrenden alınan örneklem başka aynı büyüklükteki örneklemlemler ile değiştirilebiliyorsa bu tür örneklemlemler “rastgele” olarak adlandırılır. Sabit yüzeylerde ise araştırmacının genelleme yapmak gibi bir isteği söz konusu değildir. Bir başka deyişle, seçilen koşullar dışında genelleme yapılmayacaksa ya da çalışılan evren küçük ise evren üzerinde çalışılan durumlarda sabit yüzey kullanılır. Bir yüzeyin sabit ya da rastgele alınması tamamen araştırmacının kararına bağlıdır (Atılğan, 2019; Brennan, 2001; Güler, Kaya Uyanık & Taşdelen Teker, 2012; Shavelson & Webb, 1991).

7. Genellenebilirlik ve Phi Katsayıları: G kuramında bağıl ve mutlak değerlendirme kavramlarına yukarıda değinmiştik. Sosyal bilimler alanında yapılan ölçmelerde bağıl değerlendirmeler için genellenebilirlik (G) katsayısı elde edilirken; mutlak değerlendirmeler için Phi (Φ) katsayısı elde edilir. G katsayısı, örneklemde elde edilen bireylerin gözlenen puanlarının evren puanlarına ne kadar doğru bir şekilde genellenebileceğini gösterir. G katsayısı; Eşitlik 1’de verilen formül ile elde edilir (Crocker & Algina, 2008).

$$Eb^2 = \frac{\sigma_b^2}{\sigma_b^2 + \sigma^2(\delta)} \quad (1)$$

Eb^2 : Genellenebilirlik katsayısı

σ_b^2 : Evren puanı

$\sigma^2(\delta)$: Göreli hata varyansı

Phi katsayısı güvenilirlik indeksi olarak da adlandırılır. G katsayısından farklı olarak bağıl hata varyansı $\sigma^2(\delta)$ yerine mutlak hata varyansı $\sigma^2(\Delta)$ kullanılır. Phi katsayısı; Eşitlik 2’de verilen formül ile elde edilir (Atılğan, 2019; Brennan, 2001; Shavelson & Webb, 1991).

$$\Phi = \frac{\sigma_b^2}{\sigma_b^2 + \sigma^2(\Delta)} \quad (2)$$

Φ : Phi katsayısı

σ_b^2 : Evren puanı

$\sigma^2(\Delta)$: Mutlak hata varyansı

G ve Phi katsayıları bir yüzeyli tesadüfi desenlerde klasik test kuramındaki Cronbach Alfa katsayısıyla hemen hemen aynıdır.

Tek Yüzeyli Ölçmelerde Değişkenlik Kaynakları ve Varyans Bileşenleri

Klasik test kuramında bir bireyin gözlenen puanı gerçek puanı ile hata toplamından oluşmaktadır (Crocker & Algina, 2008). G kuramında ise gözlenen puan yüzeylere bağlı olarak ifade edilmektedir. Tek yüzeyli evrende G çalışması yapılabilecek çaprazlanmış (bxm) ve yuvalanmış (m:b) olmak üzere iki desen vardır. Tek yüzeyli ölçmelerde bu desenlerdeki değişkenlik kaynakları Tablo 1.2’de gösterilmiştir.

Tablo 1.2

Tek Yüzeyli Ölçmelerde Değişkenlik Kaynakları

Değişkenlik kaynağı	Değişkenlik türü	Varyans sembolü
Birey (b)	Evren puanı	σ_b^2
Madde (m)	Hata	σ_m^2
Artık (bm,e)	Hata	$\sigma_{bm,e}^2$

“Genellenebilirlik Kuramı” 2012, Pegem Akademi, Ankara kaynağından alınmıştır.

Tablo 1.2 incelendiğinde G kuramında tek yüzeyli ölçmelerdeki değişkenlik kaynaklarının birey etkisinden (evren puan varyansı), madde etkisinden (madde güçlüğünden kaynaklı değişkenlik) ve birey-madde etkileşimi olan artık ve diğer tanımlanamayan sistematik ve tesadüfi hatalardan oluştuğu gözlenmektedir. Bu durumda tek yüzeyli çaprazlanmış bir desende bir bireyin bir madde üzerinde gözlenen puanı Eşitlik 3'te gösterilmiştir.

$$\begin{aligned}
 X_{bm} &= \mu && \text{(Genel ortalama)} && (3) \\
 &+ \mu_b - \mu && \text{(Birey etkisi)} \\
 &+ \mu_m - \mu && \text{(Madde etkisi)} \\
 &+ X_{bm} - \mu_b - \mu_m + \mu && \text{(Artık etkisi)}
 \end{aligned}$$

Eşitlik 3'te bir m maddesinde b bireyinin gözlenen puanı X_{bm} ile gösterilmiştir. Görülmektedir ki bir bireyin tek yüzeyli çaprazlanmış desende gözlenen puanı dört bileşene bağlıdır. Genel ortalama tüm bireyler için aynıdır. Birey etkisi bireyin evren puanı (μ_b) ile genel ortalama (μ) arasındaki farkı gösterir. Pozitif birey etkisi; bireyin puanının ortalamadan yüksek olduğunu gösterirken, negatif puan etkisi ise bireyin puanının ortalamadan düşük olduğu anlamına gelir. Madde etkisi ($\mu_m - \mu$) madde güçlüğüne ifade eder. Benzer şekilde pozitif madde etkisi; maddenin ortalamadan kolay, negatif madde etkisi maddenin ortalamadan zor olduğu anlamına gelir. Artık etkisi ise madde-birey etkileşiminin etkisi ile diğer tanımlanamayan sistematik ve tesadüfi hataları içerir (Atılğan, 2019; Güler, Kaya Uyanık & Taşdelen Teker, 2012).

Bir yüzeyli yuvalanmış desende bir bireyin bir maddeden gözlenen puanı (X_{mb}), bir yüzeyli çaprazlanmış desende yapılan açıklamalara benzer biçimde Eşitlik 4'te gösterilmiştir.

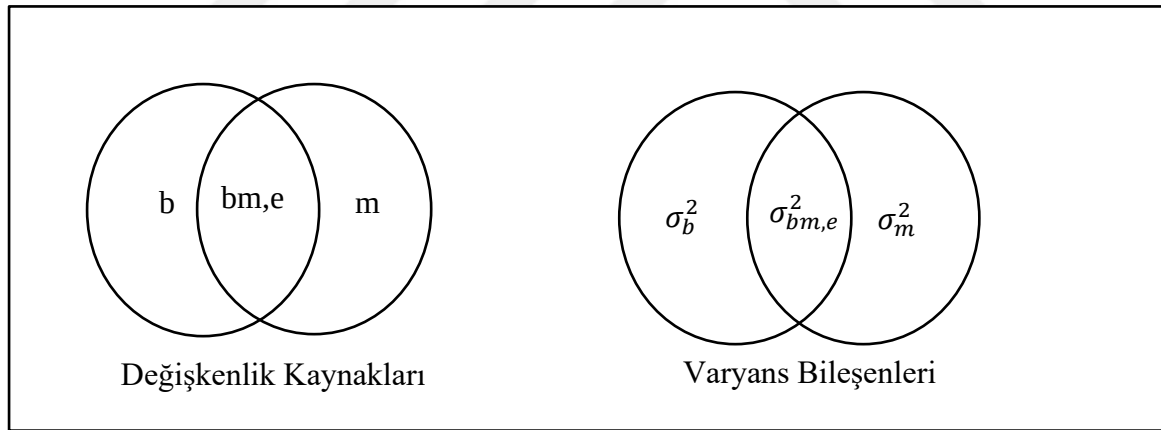
$$\begin{aligned}
 X_{mb} &= \mu && \text{(Genel ortalama)} && (4) \\
 &+ \mu_b - \mu && \text{(Birey etkisi)} \\
 &+ X_{mb} - \mu_b && \text{(Artık etkisi)}
 \end{aligned}$$

Eşitlik 4'te bir yüzeyli yuvalanmış desende bir bireyin gözlenen puanının bir yüzeyli çaprazlanmış desenden farklı olarak üç bileşene bağlı olarak hesaplandığı görülmektedir (Atılğan, 2019).

Varyans bileşenlerine bakacak olursak; tek yüzeyli çaprazlanmış bxm deseni için gözlenen puanların (X_{bm}) varyansı, üç varyans bileşeninin toplamından oluşur. Bu varyans değeri Eşitlik 5'te gösterilmiştir.

$$\sigma^2(X_{bm}) = \sigma_b^2 + \sigma_m^2 + \sigma_{bm,e}^2 \quad (5)$$

Eşitlik 5'e göre bireyin gözlenen puanlarının varyansı; birey, madde ve kalan varyanslar olmak üzere bağımsız değişkenlik kaynaklarına ayrılabilir. Bu durumu daha iyi ifade edebilmek için Şekil 1.2'de tek yüzeyli çaprazlanmış bxm deseni için değişkenlik kaynakları ve varyans bileşenleri venn diyagramı ile gösterilmiştir.



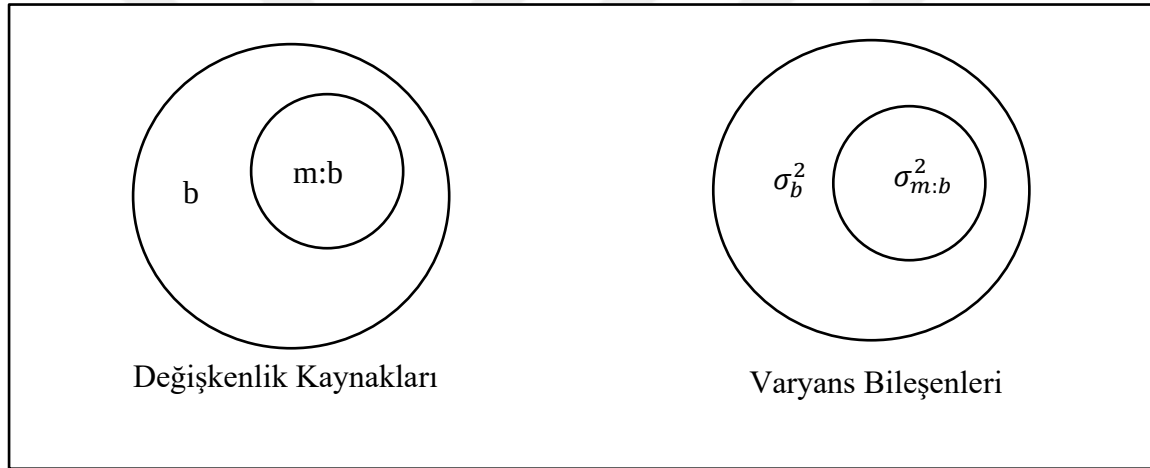
Şekil 1.2. Bir yüzeyli çaprazlanmış bxm deseni için değişkenlik kaynakları ve varyans bileşenleri. Atılğan, H. (2019). *Genellenebilirlik kuramı ve uygulaması*. Ankara: Anı.

Şekil 1.2 incelendiğinde; b ve m dairelerinin kesişim bölgesi madde ve birey yüzeylerinin ortak alanını (bm) gösterirken, aynı şekilde varyans bileşenlerinde de birey varyans bileşenleri ile madde varyans bileşenlerinin kesişim bölgesi birey-madde (bm) ortak etkisi varyansını göstermektedir (Atılğan, 2019).

Tek yüzeyli yuvalanmış m:b deseni için gözlenen puanların (X_{mb}) varyansı, iki varyans bileşeninden oluşur. Bu iki varyans bileşeninden oluşan toplam varyans Eşitlik 6'da gösterilmiştir (Atılğan, 2019).

$$\sigma^2(X_{mb}) = \sigma_b^2 + \sigma_{m,bm,e}^2 \quad (6)$$

Bir yüzeyli yuvalanmış m:b deseninde de, bir yüzeyli çaprazlanmış desende olduğu gibi; değişkenlik kaynaklarının ve varyans bileşenlerinin daha iyi anlaşılabilmesi için Şekil 1.3'ten yararlanılmıştır.

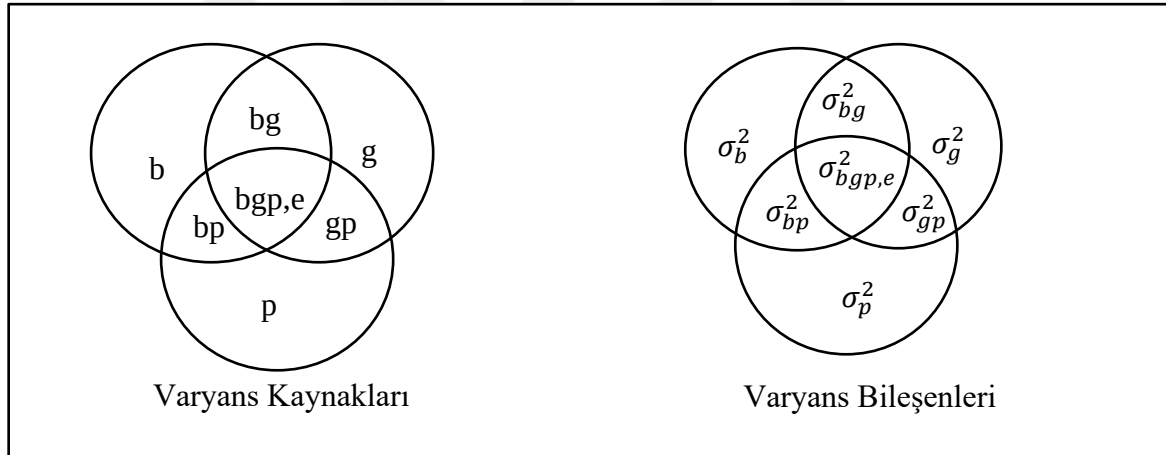


Şekil 1.3. Bir yüzeyli yuvalanmış m:b deseni için değişkenlik kaynakları ve varyans bileşenleri. Atılğan, H. (2019). *Genellenebilirlik kuramı ve uygulaması*. Ankara: Anı.

Şekil 1.3'e bakıldığında; bir yüzeyli yuvalanmış m:b deseninde iki değişkenlik kaynağı ve iki varyans bileşeni olduğu görülmektedir. Ancak unutulmamalıdır ki; buradaki m:b yuvalanma etkisi, mb ve e'yi de içermektedir. Aynı şekilde $\sigma_{m:b}^2$ varyansı da mb ve e varyanslarını içermektedir (Atılğan, 2019).

İki Yüzeyli Ölçmelerde Varyans Kaynakları ve Varyans Bileşenleri

Genellenebilirlik analizleri genel olarak iki aşamada gerçekleştirilmektedir. İlk aşamada, G çalışması ile kabul edilebilir gözlemler evreninden kestirilen varyans bileşenleri elde edilirken; ikinci aşamada, K çalışması yürütülmektedir. Bu aşamalar tek yüzeyli evrenlerde olduğu gibi iki ve çok yüzeyli evrenlerde de gerçekleştirilmektedir. Bu bölümde iki yüzeyli desenlerdeki varyans kaynaklarına ve varyans bileşenlerine değinilecektir. Bu desenlerin hepsine tek tek değinmek yerine, iki örnek desen üzerinden açıklamak daha uygun olacaktır. İlk olarak tüm bireylerin (b) her bir görevi (g) gerçekleştirdiği ve her puanlayıcının (p) her bir bireyin her bir görevini puanladığı $b \times g \times p$ tümüyle çaprazlanmış desenini ele alalım. Bu desene ait varyans kaynakları ve varyans bileşenleri Şekil 1.4'te gösterilmiştir.



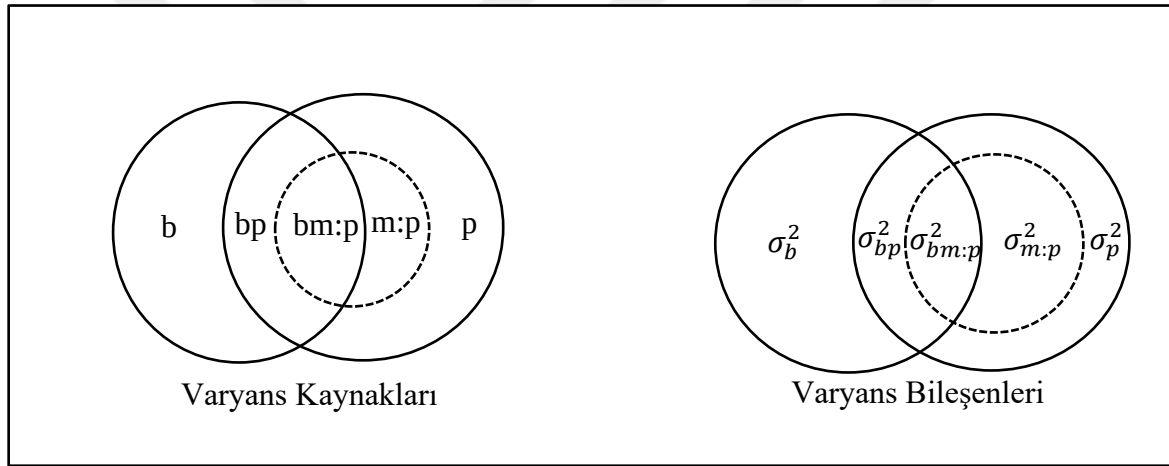
Şekil 1.4. İki yüzeyli çaprazlanmış $b \times g \times p$ deseni için varyans kaynakları ve varyans bileşenleri. Atılğan, H. (2019). *Genellenebilirlik kuramı ve uygulaması*. Ankara: Anı kaynağından alınmıştır.

Şekil 1.4 incelendiğinde iki yüzeyli tümüyle çaprazlanmış $b \times g \times p$ deseninin varyans kaynakları ve varyans bileşenleri venn şeması ile gösterilmiştir. Şekil incelendiğinde kesişim noktaları dışında kalan bölümler birey (b), görev (g) ve puanlayıcıya (p) ait varyans kaynaklarını ve varyans bileşenlerini göstermektedir. İkili kesişim noktaları ise birey-görev (bg), birey-puanlayıcı (bp) ve görev-puanlayıcı (gp) varyans kaynakları ile varyans bileşenlerine aittir. Üç dairenin de ortak kesim noktasında birey-görev-puanlayıcı ile

ölçülemeyen hata ($b_{gp,e}$) varyans kaynakları ile varyans bileşenleri yer alır. Bu durumda bireyin gözlenen puanının varyansı da bu varyans bileşenlerinin toplamına eşit olacaktır. Bireyin gözlenen puanının varyansı Eşitlik 7’de gösterilmiştir (Atılgan, 2019).

$$\sigma^2(X_{bgp}) = \sigma_b^2 + \sigma_g^2 + \sigma_p^2 + \sigma_{bg}^2 + \sigma_{bp}^2 + \sigma_{gp}^2 + \sigma_{bgp,e}^2 \quad (7)$$

İkinci olarak maddelerin (m) puanlayıcıların (p) içinde yuvalandığı, bireylerin (b) de bunlarla çaprazlandığı $bx(m:p)$ kısmi yuvalanmış desenini ele alalım. Bu desene ait varyans kaynakları ve varyans bileşenlerini Şekil 1.5’te gösterilmiştir.



Şekil 1.5. İki yüzeyli kısmi yuvalanmış $bx(m:p)$ deseni için varyans kaynakları ve varyans bileşenleri. Atılgan, H. (2019). *Genellenebilirlik kuramı ve uygulaması*. Ankara: Anı kaynağından alınmıştır.

Şekil 1.5 incelendiğinde bireylere ve puanlayıcılara ait çemberler kesişen durumda iken maddeler puanlayıcıların içinde yuvalandığından maddelere ait çember puanlayıcılara ait çemberin alt kümesi şeklinde kesikli çizgilerle gösterilmiştir. Şekle göre iki yüzeyli kısmi yuvalanmış $bx(m:p)$ deseninde beş varyans kaynağı ile beş varyans bileşeni yer almaktadır. Dolayısıyla bireyin gözlenen puanlarının varyansı bu beş varyans bileşeninin toplamına eşit olacaktır. Bu durum Eşitlik 8’de gösterilmiştir.

$$\sigma^2(X_{bnp}) = \sigma_b^2 + \sigma_p^2 + \sigma_{m:p}^2 + \sigma_{bp}^2 + \sigma_{bm:p}^2 \quad (8)$$

Araştırmanın Amacı ve Önemi

Bu araştırmada açık uçlu maddelerden oluşmuş madde takımları ile bu maddelere benzer içerikte olacak şekilde hazırlanmış olan çoktan seçmeli maddelerden oluşmuş madde takımlarının genellenebilirlik kuramı ile güvenilirlik kestirimlerinin incelenmesi amaçlanmıştır.

Son yıllarda yapılmakta olan hem ulusal sınavlarda Akademik Personel Lisansüstü Sınavı (ALES), Yabancı Dil Bilgisi Seviye Tespit Sınavı (YDS), Yükseköğretim Kurumları Yabancı Dil Sınavı (YÖKDİL) gibi; hem de uluslararası sınavlarda Yabancı Dil Olarak İngilizce Sınavı (TOEFL) ve Uluslararası Öğrenci Değerlendirme Programı (PISA) gibi madde takımlarının kullanım sıklığının arttığı görülmektedir. Puanlama avantajı sebebiyle çoğunlukla çoktan seçmeli maddelerden oluşan madde takımları kullanımı tercih edilse de açık uçlu maddelerden oluşan madde takımları kullanımı da göz ardı edilemez. Bu bağlamda çalışmada, farklı madde tipleri ile hazırlanmış birbiriyle benzer içerikte olan madde takımlarının güvenilirlik kestirimleri yapılması bu konuda çalışacak olan araştırmacılara ışık tutacaktır. Aynı zamanda bu çalışmada genellenebilirlik yaklaşımı ile güvenilirlik kestirimi yapıp bir de karar çalışması yürütülmüştür. Bu sebeple araştırmacılar hangi madde takımının kaç madde ile daha uygun olabileceğini görebilecek, yapacakları çalışmalarını kaç puanlayıcı ile yürütebileceklerine karar verebileceklerdir.

Literatür tarandığında madde takımlarının madde tepki kuramı ile ele alındığı çalışmalar (Sireci, Thissen & Wainer 1991; Wainer & Thissen, 1996) ve G kuramı ile ele alındığı çalışmalar (Lee & Frisbie, 1999; Lee, Brennan & Frisbie, 2000; Taşdelen Teker, 2014; Kaya Uyanık & Gelbal, 2018; Kaya Uyanık & Ertuna, 2022) olduğu görülmektedir. Ayrıca Gessaroli ve Folske (2002), madde takımlarını faktör analizi ile ele alırken; Kaya Uyanık ve Gelbal (2018), madde tepki modellemesinde genellenebilirlik yaklaşımı ve G kuramından elde edilen sonuçları karşılaştırmışlardır. Uluslararası alanda yapılan çalışmaların daha geniş

yer kapladığı görülmekle birlikte ulusal alanda yapılan çalışmaların sınırlı olduğu görülmüştür. Ayrıca hem açık uçlu maddelerden hem de çoktan seçmeli maddelerden oluşan testlerin karşılaştırılması yoluna gidilmemiştir. Bu açıdan bakıldığında bu araştırmanın literatüre katkı sağlayacağı düşünülmektedir. Araştırmada açık uçlu ve çoktan seçmeli maddelerden oluşan madde takımlarında madde sayısı ve madde takımı sayısının değişiminin güvenilirlik kestirimine etkisine ilişkin bir K çalışması yürütülmüştür. Ayrıca birden fazla puanlayıcının da söz konusu olduğu açık uçlu maddelerden oluşan test için puanlayıcı sayısının değişimine yönelik olarak da K çalışması gerçekleştirilmiştir. Bu nedenle araştırmanın madde takımlarının sıklıkla kullanıldığı sınavlarda kullanıcılara farklı bir öneri sunacağından önemli olduğu düşünülmektedir.

Yukarıda durumu, amacı ve önemi tanımlanan araştırmanın problem cümlesi aşağıdaki gibi ifade edilmiştir.

Problem Cümlesi

Birbirine benzer olacak şekilde hazırlanmış açık uçlu maddelerden oluşan madde takımları ile çoktan seçmeli maddelerden oluşan madde takımlarında madde takımı etkisinin ele alınmadığı ve dikkate alındığı durumlarda genellenebilirlik kuramı ile yapılan güvenilirlik kestirimleri nasıldır?

Alt Problemler

1. Açık uçlu maddelerden oluşan madde takımlarının yer aldığı başarı testlerinde;
 - a. Madde takımı etkisi ele alınmadığında birey (b), madde (m) ve puanlayıcıların (p) birbiriyle çaprazlandığı $b \times m \times p$ tümüyle çaprazlanmış desenine ait varyans bileşenleri ile G ve Phi katsayıları nedir?
 - b. Madde takımı etkisinin ele alınmadığı birey (b), madde (m) ve puanlayıcıların (p) birbiriyle çaprazlandığı $b \times m \times p$ deseninde, madde ve puanlayıcı sayısının

artırılıp azaltılmasına yönelik yapılan karar çalışmalarında G ve Phi katsayıları nasıldır?

- c. Madde takımı etkisi dikkate alınarak oluşturulan; maddelerin (m) madde takımlarında (t) yuvalandığı, bireylerin (b) ve puanlayıcıların (p) ise bunlarla çaprazlandığı $bx(m:t)xp$ deseninin varyans bileşenleri ile G ve Phi katsayıları nedir?
 - d. Madde takımı etkisinin dikkate alındığı; maddelerin (m) madde takımlarında (t) yuvalandığı, bireylerin (b) ve puanlayıcıların (p) ise bunlarla çaprazlandığı $bx(m:t)xp$ deseninde, madde takımı, madde takımı içindeki madde sayısı ve puanlayıcı sayısının artırılıp azaltılmasına yönelik yapılan karar çalışmalarında G ve Phi katsayıları nasıldır?
2. Çoktan seçmeli maddelerden oluşan madde takımlarının yer aldığı başarı testlerinde;
- a. Madde takımı etkisi ele alınmadığında bireylerin (b) maddelerle (m) çaprazlandığı bxm deseninin varyans bileşenleri ile G ve Phi katsayıları nedir?
 - b. Madde takımı etkisinin ele alınmadığı, bireylerin (b) maddelerle (m) çaprazlandığı bxm deseninde; madde sayısının artırılıp azaltılmasına yönelik yapılan karar çalışmasında G ve Phi katsayıları nasıldır?
 - c. Madde takımı etkisi dikkate alınarak oluşturulan; maddelerin (m) madde takımlarında (t) yuvalandığı, bireylerin (b) ise bunlarla çaprazlandığı $bx(m:t)$ kısmi yuvalanmış deseninin varyans bileşenleri ile G ve Phi katsayıları nedir?
 - d. Madde takımı etkisinin dikkate alındığı, maddelerin (m) madde takımlarında (t) yuvalandığı, bireylerin (b) ise bunlarla çaprazlandığı $bx(m:t)$ kısmi yuvalanmış deseninde; madde takımı sayılarının ve madde takımı içindeki madde sayılarının artırılıp azaltılması ile yapılan karar çalışmalarında G ve Phi katsayıları nasıldır?

3. Açık uçlu ve çoktan seçmeli maddelerden oluşan madde takımlarının yer aldığı başarı testlerinde madde takımı etkisinin ele alınmadığı ve dikkate alındığı durumlarda G ve Phi katsayıları nasıl değişmektedir?

Sınırlılıklar

Araştırmanın sahip olduğu sınırlılıklar aşağıda maddeler halinde sunulmuştur.

- Çalışmanın katılımcıları Ankara ilinden belirlenen bir devlet ortaokulunun 8.sınıf öğrencileri ile sınırlıdır. Çalışma grubu oluşturulurken örneklem büyüklüğünün sağlanması için Beypazarı ilçesindeki tüm ortaokullar ilk etapta dikkate alınmıştır. Fakat 6 Şubat'ta yaşanan deprem felaketinden sonra 8. sınıfların gireceği LGS sınavının konu dağılımında bir değişikliğe gidilmiştir. Öğrenciler sadece ilk dönem konularından sorumlu olduklarından ikinci dönem öğrencilerin okula devamsızlık durumlarında artış gözlenmiş ve nihai uygulama için yeteri kadar öğrenci bulunamamıştır. Bu sebeple pilot uygulama Beypazarı ilçesindeki 3 ortaokulda gerçekleştirilmiştir. Bu ortaokullar seçilirken öğrenci sayısının çokluğu göz önünde bulundurulmuştur. Nihai uygulama Ankara ili merkezinde gerçekleştirilmiştir.
- Hazırlanan başarı testleri matematik dersi ilk dönem müfredat konularından olan üslü sayılar, köklü sayılar, veri analizi ve olasılık konuları ile sınırlıdır.
- Hazırlanan başarı testleri açık uçlu ve çoktan seçmeli üçer maddelik dört madde takımı ile sınırlıdır.
- Genellenebilirlik kuramı kapsamında b_{xmxp} , $b_{x(m:t)xp}$, b_{xm} ve $b_{x(m:t)}$, desenleri ile sınırlıdır.

Sayıtlar

1. Araştırmaya katılan öğrencilerin başarı testlerine içtenlikle yanıt verdikleri varsayılmıştır. Bunun için çalışmanın amacı öğrencilere anlatılmış, kişisel bilgilerin kesinlikle bir yerde kullanılmayacağı açıklanmış ve gönüllülük esas alınmıştır.



BÖLÜM II

İLGİLİ ARAŞTIRMALAR

Bu bölümde, madde takımları ile açık uçlu veya çoktan seçmeli maddelerde genellenebilirlik kuramı üzerine yapılan araştırmalara yer verilmiştir.

Madde Takımları Üzerine Yapılan Araştırmalar

Sireci, Thissen ve Wainer (1991) çalışmalarında madde takımlarından oluşan testlerde madde takımı etkisi dikkate alınmadığında yapılan güvenilirlik kestirimlerinin farklı sonuçlar doğuracağı üzerinde durmuşlardır. Madde takımlarından oluşan testlerin güvenilirlik kestirimleri yapılırken klasik test kuramı yerine madde tepki kuramının kullanımının daha uygun olacağını belirtmişlerdir. Bu amaçla çalışmalarında 45 ikili öğeyi Bock'un (1972) nominal cevap modeli ile incelemişlerdir. Testteki 45 madde bağımsız olarak ele alındığında yapılan güvenilirlik kestirimlerinde α katsayısının ve madde tepki kuramı çerçevesinde hesaplanan $\bar{p}$ katsayısının .88 ile .86 arasında değiştiğini görmüşlerdir. Madde takımı etkisi dikkate alınarak, okuma testinin 4 okuma parçasından ve her okuma parçasının da 10 maddeden oluştuğu ele alındığında yapılan güvenilirlik kestirimlerinde α ve $\bar{p}$ katsayılarını .78 ile .76 olarak hesaplamışlardır. Maddelerin birbirine bağımlılıklarının dikkate alındığı durumda güvenilirlik katsayılarının düştüğünü görmüşlerdir. Sonuç olarak; eğer maddeler

yerel madde bağımsızlığına sahiplerse madde tabanlı güvenilirlik belirleme yöntemi kullanılabileceğini, fakat maddeler yerel bağımlı ise madde takımı etkisinin dikkate alınarak güvenilirlik kestirimleri yapılmasını önermektedirler.

Wainer ve Thissen (1996), çalışmalarında güvenilirlik katsayılarındaki küçük düşüşlerin sonuçlarının neler olduğunu, yerel madde bağımlılığının güvenilirliği nasıl azaltacağını ve madde takımlarından oluşan okuma parçalarının güvenilirliği nasıl etkileyeceği üzerinde durmuşlardır. Madde takımlarında maddeler arasında yerel madde bağımlılığı olacağından güvenilirliğin düşeceğini belirtmişlerdir. Ayrıca madde takımlarının uzunlukları arttıkça da maddeler arası yerel madde bağımlılığı artacağından güvenilirliğin düşeceği sonucuna ulaşmışlardır. Madde takımı içeren üç testin güvenilirliklerini karşılaştırdıklarında testi cevaplamak için verilen sürenin fazla olduğu durumlarda, güvenilirliğin düşük olmasına karşın madde takımlarının kullanılmasının daha uygun olacağı sonucuna varmışlardır.

Lee ve Frisbie (1999), yaptıkları çalışmada madde takımlarından oluşan test puanlarının güvenilirliğini Genellenebilirlik teorisi modeli ile kestirmeye çalışmışlardır. Uygulamayı farklı alanlarda, farklı sınıf seviyelerinden seçilmiş bir çalışma grubu üzerinde gerçekleştirmişlerdir. Madde takımlarını oluşturan madde sayıları da farklılık göstermektedir. Cronbach Alfa yöntemi ile madde takımı etkisi dikkate alınmadan kestirilen güvenilirliğin Genellenebilirlik teorisi ile madde takımı etkisi dikkate alındığında kestirilen güvenilirlikten %4 daha yüksek olduğunu görmüşlerdir. Ayrıca genellenebilirlik katsayısına madde takımlarının içindeki madde sayısından ziyade madde takımlarının sayısının daha fazla etki ettiği sonucuna ulaşmışlardır.

Lee, Brennan ve Frisbie (2000) çalışmalarına farklı sayıda maddelerden oluşan madde takımlarının test puanı üzerindeki etkisini dahil etmişlerdir. Öncelikle madde takımı etkisi dikkate alınmadığında elde edilen güvenilirlik katsayısını Cronbach Alfa yöntemi ile incelemişlerdir. Elde edilen bu değer .885 olarak bulunmuştur. İkinci aşamada test formu üç bölümde ele alınmıştır. Bu durumda elde edilen güvenilirlik katsayıları Cronbach Alfa yöntemi ile madde takımlarının toplam puanları için .598, madde takımlarının ortalama

puanları için .684 olarak hesaplanmıştır. Üçüncü aşamada madde takımlarındaki madde sayıları dikkate alınarak genellenebilirlik katsayısı .716 olarak bulunmuştur. Son aşamada ise testin üç kategoriden oluştuğu varsayılarak tabakalı alfa katsayısı .941 olarak hesaplanmıştır.

Gessaroli ve Folske (2002), madde takımlarından oluşan testlerin güvenilirliği çalışmasında madde takımlarını sabit veya rastgele alarak faktör analizi ile güvenilirliği kestirmeyi amaçlamışlardır. Temel olarak hiyerarşik faktör analizi kullanılan bu çalışmada Cronbach alpha katsayısından daha üstün bir güvenilirlik kestirimi yapıldığı belirtilmiştir.

Taşdelen Teker (2014), madde takımlarının güvenilirlik ve değişen madde fonksiyonu üzerine etkisi adlı doktora tezinde; farklı madde sayıları içeren madde takımlarının güvenilirlik kestirimini dengelenmemiş desenlerde genellenebilirlik kuramı (GK) ile incelemiştir. Çalışmada, madde takımlarından oluşan çoktan seçmeli 20 maddelik bir İngilizce testini yanıtlayan 1500 kişilik bir veri grubu kullanılmıştır. Madde takımı etkisi dikkate alınmadığında kestirilen güvenilirlik katsayısının, madde takımı etkisi dikkate alındığında kestirilenden daha yüksek olduğu sonucuna ulaşmıştır. Ayrıca güvenilirliği artırmak için madde takımı sayısındaki artışın, madde takımını oluşturan madde sayısındaki artıştan daha etkili olduğunu vurgulamıştır. Diğer yandan ise, madde takımlarını oluşturan madde sayılarının farklı olmasının güvenilirliği olumsuz etkilediğini görmüştür.

Kaya Uyanık ve Gelbal (2018), çalışmalarında simülasyon verisi kullanarak madde tepki modellemesinde genellenebilirlik yaklaşımı (MTMG) ile iki yüzeyli desenler üzerinde çalışmışlar ve elde edilen sonuçları genellenebilirlik kuramı (GK) ile elde edilen sonuçlar ile karşılaştırmışlardır. Bu bağlamda simülasyonda üretilen madde takımları verileri etki, uzunluk ve sayı açısından farklılıklar göstermektedir. Madde takımı uzunluğu 6 ve 9, madde takım sayıları ise 3 ve 5 olup kullanılan madde sayıları 18, 30, 27 ve 45'tir. Araştırma sonucuna göre kestirilen varyans değerlerinin MTMG yaklaşımı ve GK ile elde edilen sonuçlar arasında farklılık olmadığı gözlemlenmiştir. Fakat MTMG yaklaşımı ile hata

varyansının etkileşim varyansından ayrı olarak gözlenebileceğine ulaşılmıştır. Ayrıca birey madde etkileşiminin güvenilirlik ile ters orantılı olduğu yani etkileşim küçük olduğunda güvenilirlik yüksek, etkileşim büyük olduğunda güvenilirliğin düşük olduğu gözlemlenmiştir. Madde takımı sayısı açısından bakıldığında ise, madde sayısı artışının güvenilirliği olumlu bir şekilde etkilediği sonucuna ulaşılmıştır.

Kaya Uyanık ve Ertuna (2022), çalışmalarında açık uçlu maddelerden oluşan bir testte madde takımı etkisini ele almışlardır. Çalışmanın amacı açık uçlu maddelerden oluşan madde takımlarının etkisini değerlendirmek ve bu etki göz ardı edildiğinde olası hataları tespit etmektir. Bu amaçla yapılan uygulama 380 öğretmen adayı üzerinde gerçekleştirilmiştir. Öğretmen adaylarının her birinin iki açık uçlu maddeden oluşan iki madde takımına verdikleri yanıtları üç puanlayıcı değerlendirmiştir. Çalışmalarında madde takımı etkisinin göz ardı edildiği tüm bireylerin(b) tüm maddeleri(m) aldığı ve tüm puanlayıcılar(p) tarafından değerlendirildiği $b \times p \times m$ tümüyle çaprazlanmış desen ile madde takımı(t) etkisinin dikkate alındığı $b \times p \times (m:t)$ yuvalanmış desen oluşturulmuş olup oluşturulan bu modeller genellenebilirlik kuramı ile analiz edilmiştir. Analiz sonuçları, puanlayıcı etkisinin her iki modelde de hafif bir etkiye sahip olduğunu ancak madde takımı etkisi göz ardı edildiğinde kestirilen güvenilirliğin daha yüksek olduğunu göstermektedir.

Açık Uçlu veya Çoktan Seçmeli Maddelerde Genellenebilirlik Kuramı Üzerine

Yapılan Araştırmalar

Güler ve Gelbal (2010), açık uçlu matematik maddelerinin güvenilirliğini hem klasik test kuramına hem de genellenebilirlik kuramına göre incelemişlerdir. Çalışmalarında; TIMSS-1999'da yer alan açık uçlu 24 matematik maddesini, 2007 yılında 203 öğrenciye uygulamışlardır. Klasik test kuramına göre elde edilen puanların iç tutarlılığını 0,92 olarak kestirirken, genellenebilirlik kuramına göre elde edilen puanların G katsayısını 0,92 ve Phi katsayısını 0,90 olarak hesaplamışlardır. Çalışmalarında dört puanlayıcının puan

ortalamaları arasında az bir fark olmasına rağmen uyumlu puanlama yaptıklarını gözlemlemişlerdir.

Çakıcı Eser ve Gelbal (2013) genellenebilirlik kuramı ve lojistik regresyona dayalı hesaplanan puanlayıcılar arası tutarlılığı karşılaştırmışlardır. Araştırmalarında 106 öğrenciye 15 açık uçlu ve kısa cevaplı maddelerden oluşan bir ölçme aracı verilmiştir. Öğrencilerin vermiş oldukları yanıtlar üç puanlayıcı tarafından dereceli puanlama anahtarı kullanılarak puanlanmıştır. Elde edilen veriler G kuramı ve LR analizi ile madde bazında ve testin tamamına dayalı olarak incelenmiştir. Çalışma sonucunda puanlayıcılar arası güvenilirliği belirlemede G kuramı ile LR analizinin sonuçlarının paralel olduğuna ulaşılmıştır. Ancak G kuramının LR analizine göre daha hassas sonuçlar verdiği gözlenmiştir.

Güler ve Taşdelen Teker (2015) çalışmalarında, açık uçlu maddelerde dört farklı yaklaşımla ele aldıkları puanlayıcılar arası güvenilirliği değerlendirmişlerdir. Buna göre 43 öğrencinin 10 açık uçlu maddeye vermiş oldukları yanıtlar iki puanlayıcı tarafından puanlanmıştır. Elde edilen veriler korelasyon, ortalamaların karşılaştırılması, uyuşma yüzdesi ve genellenebilirlik kuramı ile incelenmiştir. Çalışmalarında elde edilen puanlayıcılar arası güvenilirlik kestirimlerinde en büyük değerin 0,90 ile G kuramından elde ettikleri değer olduğu sonucuna ulaşmışlardır. Ayrıca G kuramının pek çok hata kaynağını aynı anda ele almasından dolayı diğer yöntemlere göre özellikle puanlayıcılar arası güvenirliliğin belirlenmesinde kullanılmasını önermişlerdir.

Kamış ve Doğan (2017), yaptıkları çalışmada genellenebilirlik kuramında karar çalışmalarından elde ettikleri G ve Phi katsayılarını gerçek durumlardan elde ettikleri G ve Phi katsayıları ile karşılaştırmışlardır. Geliştirdikleri dört adet açık uçlu maddeden oluşan başarı testini lisans öğrencileri üzerinde uygulamışlardır. Uygulama sonucunda elde ettikleri G ve Phi katsayıları ile karar çalışmalarından elde ettikleri G ve Phi katsayılarının birbirine yakın ama farklı sonuçlar olduğuna ulaşmışlardır. Eğer evrenden yansız olarak puanlayıcı seçimi yapılamayacaksa bu katsayıların farklılık gösterebileceğinin göz önünde bulundurulması gerektiğine değinmişlerdir.

Başol ve Şevran (2019), farklı ölçme araçlarının güvenilirliğini genellenebilirlik kuramı yardımı ile incelemişlerdir. Çalışmalarında çoktan seçmeli test, yapılandırılmış grid ve performans görevi olmak üzere üç farklı ölçme aracının güvenilirlik katsayılarını G ve K çalışmaları yürüterek karşılaştırmışlardır. Araştırmanın çalışma grubunu; 2014-2015 eğitim öğretim yılında Antalya’da öğrenim gören iki farklı okuldan rastgele seçilmiş 23’ü kız, 17’si erkek olan toplam 40 altıncı sınıf öğrencisi oluşturmaktadır. Araştırmacılar tarafından geliştirilen ölçme araçları öğrencilere uygulanmıştır. Elde edilen sonuçlara göre, çoktan seçmeli test G katsayısı .68 ve Phi katsayısı .64 ile son sırada yer almıştır. Yapılandırılmış gride ait G katsayısı .91 ve Phi katsayısı .83 olarak hesaplamıştır. Son olarak performans görevinin G katsayısı .89 ve Phi katsayısı .86 şeklinde bulunmuştur. Yaptıkları K çalışmalarında da bu üç ölçme aracına ait G ve Phi katsayılarının değişimin paralel olduğunu görmüşlerdir.

Akın Arıkan ve Gümüş Özyıldırım (2020), çoklu gösterimlerin kullanıldığı matematik problemlerine ait çözümleri puanlayıcı güvenilirliği açısından incelemişlerdir. Araştırmanın çalışma grubunu 8. sınıfta öğrenim gören 54 öğrenci oluşturmuştur. Bu öğrencilere dört farklı gösterimde açık uçlu maddeler şeklinde hazırlanan matematik problemleri sunulmuş ve çözümleri dereceli puanlama anahtarı kullanılarak 5 puanlayıcı tarafından puanlanmıştır. Çalışmanın verileri G kuramı kapsamında oxgxp (öğrenci x görev x puanlayıcı) tümüyle çaprazlanmış desen kullanılarak ele alınmıştır. Araştırma sonunda farklı gösterimdeki problemlerin puanlayıcı güvenilirliğinde etkili olduğu sonucuna ulaşmışlardır.

Demirbilek Zorba (2020) yüksek lisans tezinde, açık uçlu maddelerden oluşan personel alım sınavı sonuçlarını genellenebilirlik kuramı kapsamında ele almıştır. Teste katılan 211 bireye dört açık uçlu madde sorulmuş ve bu maddelerden iki tanesini yapmaları istenmiştir. Bireylerin en çok seçtikleri maddeler belirlenerek veri sınırlandırılmış ve 93 bireyin verileri iki puanlayıcı tarafından puanlanmıştır. Elde edilen bu veriler bmxpx (birey-madde-puanlayıcı) tümüyle çaprazlanmış desen ile ele alınmıştır. 40 bireyden elde edilen veriler üzerinde de bx(m:p) yuvalanmış deseni incelenmiştir. Çalışma sonucunda bmxpx desenine

ait G ve Phi katsayıları sırasıyla 0,33 ve 0,29; $bx(m:p)$ desenine ait G ve Phi katsayıları sırasıyla 0,76 ve 0,64 olarak hesaplanmıştır. K çalışmaları sonucunda; $bmxp$ deseninde puanlayıcı sayısının artırılması, $bx(m:p)$ deseninde ise madde sayısının artırılması ile güvenilirliğin arttığı sonucuna ulaşmıştır.

Taştan ve Uzun (2021), çalışmalarında genellenebilirlik kuramında çok yüzeyli desenleri ele almışlardır. Bu bağlamda, öğrencilerin geometri performans görevi puanlarının birden fazla puanlayıcı tarafından puanlanmasıyla oluşturulan desenlerden elde edilen G çalışmalarının sonuçlarını incelemişlerdir. Araştırmanın çalışma grubunu, 2014-2015 eğitim öğretim yılında Mersin ilinde öğrenim gören 128 12. sınıf öğrencisi oluşturmaktadır. Sekiz puanlayıcı ve sekiz açık uçlu geometri maddelerinden oluşan görevlerin yer aldığı çalışmada; $(b:o)gxpx$, $(b:o)x(g:p)$ ve $(b:(p:o))xg$ desenleri (b: öğrenci, o: okul, g: görev ve p: puanlayıcı) için G çalışması yürütülmüştür. Aynı veri seti üzerinden kurgulanan bu desenlerde varyans bileşenlerine ait büyüklüklerin farklılıklar gösterdiğine ulaşmışlardır. Ayrıca artık varyans bileşenin de değiştiğini gözlemlemişlerdir. Bu durumda; G kuramında seçilecek olan desenin doğruluğunun, güvenilirlik ve geçerlik yorumlarını etkileyeceği sonucuna ulaşmışlardır. Çalışmalarında $(b:(p:o))xg$ deseninin varyans bileşenlerinin açıklama oranının az olmasına rağmen G ve Phi katsayılarının diğer desenlere göre daha yüksek olduğunu görmüşlerdir.

BÖLÜM III

YÖNTEM

Bu bölümde araştırmanın yöntemi, çalışma grubu, veri toplama aracı, uygulama süreci ve verilerin analizine yer verilmiştir.

Araştırmanın Yöntemi

Bu araştırmada çoktan seçmeli ve açık uçlu maddelerden oluşan madde takımlarının yer aldığı başarı testlerinde madde takımı etkisinin ele alınmadığı ve dikkate alındığı durumlarda genellenebilirlik kuramı çerçevesinde G ve Phi güvenirlik katsayılarının elde edilmesi ve karşılaştırılması amaçlanmıştır. Araştırma var olan kuramın farklı durumlarda sınanması ile yeni bilgi edinmeye yönelik olduğundan *temel* bir araştırmadır (Karasar, 1994).

Çalışma Grubu

Araştırmanın çalışma grubunu 2022-2023 eğitim öğretim yılında Ankara ili Beypazarı ilçesi Milli Eğitim Bakanlığı'na bağlı ortaokullarda öğrenim gören sekizinci sınıf öğrencileri oluşturmaktadır. Çalışma grubu oluşturulurken örneklem büyüklüğünün sağlanması için Beypazarı ilçesindeki tüm ortaokullar ilk etapta dikkate alınmıştır. Fakat 6 Şubat'ta yaşadığımız deprem felaketinden sonra 8. sınıfların gireceği LGS sınavının konu

dağılımında bir değişikliğe gidilmiştir. Öğrenciler sadece ilk dönem konularından sorumlu olduklarından ikinci dönem öğrencilerin okula devamsızlık durumlarında artış gözlenmiş ve uygulama için yeteri kadar öğrenci bulunamamıştır. Bu sebeple pilot uygulama Beypazarı ilçesindeki 3 ortaokulda gerçekleştirilmiştir. Bu ortaokullar seçilirken öğrenci sayısının çokluğu göz önünde bulundurulmuştur. Nihai uygulama ise Beypazarı ilçesindeki öğrenci sayısı yetersiz kaldığından Ankara ili merkez ilçesindeki bir devlet ortaokulunda yapılmıştır. Bu ortaokul belirlenirken pilot uygulamadaki ortaokulların başarı düzeylerine benzer olacak şekilde seçilmesine dikkat edilmiştir.

Pilot uygulamada 119 öğrenci açık uçlu maddelerden oluşan matematik başarı testini, 115 öğrenci ise çoktan seçmeli maddelerden oluşan matematik başarı testini çözmüşlerdir. Açık uçlu maddelerden oluşan başarı testini çözen 119 öğrenciden 31'i ya maddelerin tamamını boş bıraktığı ya da sıfır puan aldığı için analizde yer almamıştır. Sonuç olarak pilot uygulamanın analizinde; 88 öğrencinin açık uçlu maddelere verdikleri yanıtlar, 115 öğrencinin de çoktan seçmeli maddelere vermiş oldukları yanıtlar incelenmiştir.

Nihai uygulamada 198 öğrenci çoktan seçmeli maddelerden oluşan başarı testini çözerken 228 öğrenci de açık uçlu maddelerden oluşan başarı testini cevaplandırmıştır. Başarı testini çözen öğrenciler incelendiğinde, bu öğrencilerin 157 tanesinin hem açık uçlu hem de çoktan seçmeli maddelerden oluşan başarı testlerini çözdükleri görülmüştür. Diğer öğrencilerin ise uygulamanın yapıldığı günlerde devamsızlık yapmalarından dolayı bu testlerden sadece birinde yer aldıkları tespit edilmiştir. Uygulamadaki amacımız aynı bireylerin hem çoktan seçmeli maddelerden hem de açık uçlu maddelerden oluşan başarı testlerindeki durumlarını incelemek olduğundan 112 öğrencinin test sonuçları çalışmamızda yer almamıştır. Ayrıca bu 157 öğrencinin de 29 tanesinin açık uçlu maddelerden oluşan başarı testini ya hiç yanıtlamadıkları ya da yanlış yanıtlayıp sıfır puan aldıkları gözlenmiştir. Bu sebeple analizde bu öğrencilerin sonuçlarına yer verilmemiştir. Nihai uygulamada 128 öğrencinin hem açık uçlu hem de çoktan seçmeli maddelerden oluşan başarı testlerine verdikleri yanıtlar değerlendirilmiştir.

Veri Toplama Aracı

Bu araştırma için gerekli olan veriler hazırlanan iki ayrı matematik başarı testi ile toplanmıştır. Öncelikle pilot uygulama için benzer içerikte olan bir tanesinde açık uçlu maddelerin yer aldığı madde takımlarından, diğerinde ise çoktan seçmeli maddelerin yer aldığı madde takımlarından oluşan iki adet matematik başarı testi hazırlanmıştır. Bu testlerin her birinde dört madde takımı ve her takımda dört madde yer almıştır. Hazırlanan bu maddeler 2022-2023 eğitim öğretim yılı 8.sınıf müfredatı ilk dönem konularından bazılarını kapsamaktadır. 2022-2023 eğitim öğretim yılı 8.sınıf matematik dersi müfredatı göz önüne alındığında hazırlanan maddelerin “üslü ifadeler, karaköklü ifadeler, veri analizi ve basit olayların olma olasılığı” konularından birini içerdiği görülmektedir.

Açık uçlu maddelerden oluşan madde takımlarının değerlendirilmesinde rubrik olarak da bilinen dereceli puanlama anahtarından (DPA) yararlanılmıştır. DPA; bir çalışma, ödev ya da performans için neyin ne derece önemli olduğunu; başka bir deyişle öğrencinin verdiği yanıttan ya da gösterdiği performanstan ne kadar puan alacağını kademeli olarak gösteren değerlendirme ölçeğidir. Farklı biçimlerde ele alınabilmekle birlikte tüm DPA’larda önemli olan iki ortak noktanın olduğu görülmektedir. Bunlar ölçüt ve kademeli puanlamadır (Goodrich Andrade, 1996, 2000). Araştırma kapsamında açık uçlu her bir madde için ayrı bir DPA ve değerlendirme formu hazırlanmıştır. Buna göre tam doğru yanıtlar 3 puan üzerinden; kısmen doğru yanıtlar ise kendi içerisinde kademelendirilmiş olup 2 ve 1 puan üzerinden değerlendirmeye tabi tutulmuştur. Boş ve diğer yanıtlar ise 0 puan olarak değerlendirmeye alınmıştır.

Analizlerde daha güvenilir sonuçlar elde etmek amacıyla üç matematik öğretmeni puanlayıcı olarak görev almıştır. Matematik öğretmenleri puanlamalarını birbirinden bağımsız olarak yürütmüşlerdir. Önce bir öğretmen puanlamasını yaptıktan sonra diğeri puanlamasını gerçekleştirmiş ve verdikleri puanları görmemişlerdir. Puanlayıcılara, maddeler ve DPA hakkında puanlama öncesi kısa bir açıklama yapılmıştır.

Geliştirilmiş açık uçlu ve çoktan seçmeli testlerin pilot formları için 7 uzmandan görüş alınmıştır. Bu bağlamda görüşlerine başvurulmuş uzmanlardan üçü hem matematik hem de ölçme ve değerlendirme alanında uzman olup biri doçent düzeyinde öğretim üyesi diğer ikisi ise yüksek lisans düzeyinde öğretmendir. Ayrıca bir lisans mezunu matematik öğretmeninden görüş alınmıştır. Ölçme ve değerlendirme alanında doçent olan iki uzmandan görüş alınmıştır. Böylece hazırlanmış olan başarı testlerinin uygun olup olmadığına karar verilmiştir. Uzman görüşleri dikkate alınarak testlerin son hali hazırlanmıştır. Ayrıca başarı testlerinin son hali elde edilince lisans düzeyinde bir türkçe öğretmenin yazım ve imla konusunda bilgisine başvurulmuştur. Uzmanlar görüş bildirirken her bir madde için “uygundur” ya da “uygun değildir” ifadelerinden uygun olan birini seçmişlerdir. Uzmanlar aynı madde için aynı seçeneği işaretlerse görüş birliği, farklı seçeneği işaretlerse görüş ayrılığı olarak kabul edilmiştir. Bu çalışmada, geçerlik ve güvenilirliğin sağlanması için uzmanlar arası görüş birliği hesaplanmıştır. Bu amaçla, Miles ve Huberman’ın (1994) güvenilirlik formülü kullanılarak uzmanlar arası uyum yüzdelere bakılmıştır. Formüle göre uyum yüzdesi “Güvenirlik = (Görüş Birliği / Görüş Birliği+ Görüş Ayrılığı) * 100” şeklinde ifade edilmektedir. Bu doğrultuda alınan uzman görüşlerine göre uzmanlar arası uyum yüzdeleri Miles ve Huberman’ın (1994) formülünden yararlanılarak hesaplanmış ve uzmanlar arası uyum yüzdeleri ortalaması %85 olarak bulunmuştur. Araştırmanın güvenilir kabul edilebilmesi için, güvenilirlik hesaplarının %70’in üzerinde olması gerekmektedir (Miles & Huberman, 1994). Dolayısıyla bu çalışmada elde edilen sonuç, puanlayıcılar arası uyumun sağlandığını belirtmektedir.

Uygulamadaki amaç açık uçlu ve çoktan seçmeli maddelerden oluşan madde takımlarının benzer içerikte olması ve bu testlerin güvenilirliğinin incelenmesi olduğundan, uzmanlardan test maddelerinin benzer içerikte olup olmadıklarına dair de görüş bildirmeleri istenmiştir. Bu bağlamda uzmanlar “benzer-kısmen benzer-benzer içerik değil” ifadeleri doğrultusunda görüş bildirmişlerdir. Gelen uzman görüşleri doğrultusunda elde edilen verilerin uyum yüzdeleri hesaplanmış ve bu yüzdelere aşağıda Tablo 3.1’de gösterilmiştir.

Tablo 3.1

Maddelerin İçerik Yönünden Benzerliğine Dair Uyum Yüzdeleri

Madde	“Benzer içerik” diyen kişi sayısı	“Kısmen-Benzer içerik” diyen kişi sayısı	“Benzer içerik değil” diyen kişi sayısı	Uyum Yüzdesi
1	4	2	0	83
2	6	0	0	100
3	6	0	0	100
4	4	2	0	83
5	5	1	0	92
6	5	1	0	75
7	5	1	0	92
8	5	1	0	92
9	4	2	0	83
10	6	0	0	100
11	6	0	0	100
12	4	2	0	83
13	4	2	0	83
14	3	3	0	75
15	4	2	0	83
16	6	0	0	100

Tablo 3.1 incelendiğinde uyum yüzdesi değerlerinin %70’in üzerinde olduğu görülmektedir.

Bu sebeple maddelerin benzer içerikte olduklarına karar verilmiştir. Açık uçlu ve çoktan seçmeli maddelerin tam olarak paralel olmasının sağlanmasının çok zor olmasından dolayı, madde yazımında maksimum eşdeğerlik göz önüne alınmıştır.

Pilot uygulama çalışmasının amacı doğrultusunda madde güçlük ve madde ayırt edicilik indekslerine bakılmıştır. Çoktan seçmeli maddelerden oluşan başarı testine ait maddelerin madde güçlük ve ayırt edicilik indeksleri TAP programı ile incelenmiş ve elde edilen veriler Ek 1’de gösterilmiştir. Açık uçlu maddelerden oluşan başarı testine ait maddelerin madde güçlük ve ayırt edicilik indeksleri Excel programı yardımı ile elde hesaplanmıştır.

Açık uçlu maddelerin madde ayırt edicilik (r) indeksleri hesaplanırken alt-üst grup yöntemi kullanılmıştır. Buna göre öğrenci puanları en yüksekten en düşüğe göre sıralanmış ve en başarılı %27 ile en başarısız %27'lik grupların verileri dikkate alınarak maddelerin ayırt ediciliği hesaplanmıştır. Maddelerin madde güçlüğü hesaplanırken tüm grup verileri dikkate alınmıştır. Hesaplamalar yapılırken Eşitlik 9 ve Eşitlik 10'da yer alan formüllerden (Şata, 2022) yararlanılmıştır.

$$p = \frac{\text{Maddeden alınan ortalama puan} - \text{Maddeye verilen en düşük puan}}{\text{Maddeye verilen en yüksek puan} - \text{Maddeye verilen en düşük puan}} \quad (9)$$

$$r = \frac{\text{Üst grubun maddedeki ortalaması} - \text{Alt grubun maddedeki ortalaması}}{\text{Maddeye verilen en yüksek puan} - \text{Maddeye verilen en düşük puan}} \quad (10)$$

Çalışmada uygulanan çoktan seçmeli ve açık uçlu maddelerden oluşan başarı testlerine ait istatistiklere Tablo 3.2'de yer verilmiştir.

Tablo 3.2

Pilot Uygulamadaki Testlerin Madde İstatistikleri

Madde Takımı	Madde No	Çoktan seçmeli maddelere ait madde istatistikleri		Açık uçlu maddelere ait madde istatistikleri	
		p	r	p	r
1	1	0,50	0,47	0,36	0,32
	2	0,46	0,46	0,33	0,71
	3	0,43	0,38	0,32	0,62
	4	0,33	0,24	0,17	0,40
2	5	0,58	0,62	0,47	0,84
	6	0,63	0,52	0,34	0,67
	7	0,37	0,57	0,09	0,22
	8	0,34	0,37	0,01	0,05
3	9	0,42	0,43	0,52	0,60
	10	0,18	0,19	0,03	0,10
	11	0,15	0,14	0	0
	12	0,33	0,45	0	0
4	13	0,48	0,48	0,14	0,41
	14	0,42	0,60	0,06	0,22
	15	0,36	0,55	0,08	0,27
	16	0,29	0,53	0,08	0,21

Tablo 3.2’deki analiz sonuçlarında yer alan madde güçlükleri (p) incelendiğinde çoktan seçmeli maddelerin öğrencilere daha kolay geldiği görülürken açık uçlu maddelerde ise çok zorlandıkları görülmektedir. Bunun sebebi ise öğrencilerin açık uçlu maddelere karşı daha ön yargılı olduğu veya çoktan seçmeli testlere daha aşina olduğu söylenebilir. Ayrıca çoktan seçmeli maddelerde şans başarısı da göz önünde bulundurulduğunda öğrenci tam olarak doğru yanıt bulamamışsa bile yakın olan bir seçeneği işaretlemiş olabilir. Açık uçlu maddelerde ise böyle bir durum mümkün olmadığından öğrencilerin çoğunlukla maddeleri boş bıraktıkları görülmektedir. Özellikle üçüncü madde takımına ait olan 11. ve 12. maddeler hiçbir öğrenci tarafından yanıtlanmadığından bu maddelere ait madde istatistikleri sıfır olarak çıkmıştır. Bu durumun sebebinin de bu madde takımına ait maddelerin yanıtlarının serbest bırakılmış yanıtlar olarak hazırlanmış olduğu söylenebilir. Çünkü diğer madde takımlarındaki maddelerin yanıtları sınırlandırılmış iken, bu madde takımına ait maddelerin yanıtları serbest bırakılmış olarak hazırlanmıştır. Bu maddelerde öğrenciler maddeleri yanıtlama ile uğraşmak yerine boş bırakmayı tercih etmiş olabilirler. Sonuçların bu şekilde farklılık arz etmesinden dolayı ölçme ve değerlendirme alanında iki öğretim üyesinden de görüş alınarak madde istatistiklerini birebir karşılaştırmak yerine maddelerin benzer içerikte olduğuna dair uzman görüşlerinin de kullanılmasına karar verilmiştir. Ayrıca nihai uygulamada kullanılacak olan maddelere karar verilirken çoktan seçmeli maddelerin ayırt edicilikleri (r) göz önüne alınmış ve bu doğrultuda bir elemeye gidilmiştir.

Tablo 3.2’de yer alan madde ayırt ediciliklerine baktığımızda birçok maddenin iyi düzeyde ayırt edici olduğunu görmekteyiz. Genellikle, madde ayırtıcılığı 0,20 ile 0,30 arasında olan maddeler testte kullanılabilir nitelikte; 0,30 ile 0,40 arasında olan maddeler iyi; 0,40’tan daha yüksek olan maddeler ise çok iyi sayılmaktadır. Ayırtıcılığı 0,20’den daha düşük maddelerin düzeltilip geliştirilerek kullanılması önerilir (Özçelik, 2013, s.134-136). Bu durum dikkate alındığında bu araştırmada dengelenmiş desenler incelendiğinden dolayı her takımdan bir maddenin çıkarılmasına karar verilmiştir. Bu maddeler ayırt edicilikleri düşük olan 4, 8, 11 ve 16. maddelerdir. Ayrıca 10. maddenin de düzeltilerek teste alınmasına karar verilmiştir.

Bu bağlamda madde de düzeltmeye gidilmiş ve nihai uygulama için oluşturulan başarı testlerinde 4'er takım ve her takımda 3'er madde yer almıştır.

Pilot uygulama ile amaç; hazırlanan açık uçlu ve çoktan seçmeli test maddelerinin madde güçlük ve ayırt edicilik indekslerini belirlemek olmuştur. Bu amaç doğrultusunda hangi maddelerin çıkarılması veya düzenlenmesi gerektiğine karar verilerek nihai test oluşturulmuştur. Çoktan seçmeli maddelerden oluşan başarı testinin nihai formuna Ek 2'de, açık uçlu maddelerden oluşan başarı testinin nihai formuna Ek 3'te ve açık uçlu maddeler için hazırlanan DPA'ya Ek 4'te yer verilmiştir.

Uygulama Süreci

Hazırlanan başarı testleri 8.sınıf birinci dönem konularını içerdiğinden pilot uygulama bu konular işlendikten sonra gerçekleştirilmiştir. Pilot uygulama gerçekleştirildikten sonra gerekli analizler yapılmış ve nihai testler oluşturulmuştur.

Nihai uygulama için bir gruba ilk olarak açık uçlu maddelerden oluşmuş başarı testi uygulanırken diğer gruba çoktan seçmeli maddelerden oluşmuş başarı testi uygulanmıştır. Böylece farklı madde formatlarından oluşan testlerin uygulanma sırasının bir etki yaratmasının önüne geçilmek istenmiştir. İkinci testlerin uygulanması için on gün ara verilmiştir. 10 gün sonra ise önce açık uçlu maddelerden oluşan testin uygulandığı gruba çoktan seçmeli maddelerden oluşan test, çoktan seçmeli maddelerden oluşan testin uygulandığı gruba ise açık uçlu maddelerden oluşan test uygulanmıştır.

Verilerin Analizi

Verilerin analizi iki aşamada gerçekleştirilmiştir. Öncelikle pilot uygulama sonucu elde edilen madde istatistikleri hesaplanmış, daha sonra ise nihai uygulama sonucu elde edilen veriler araştırma alt problemleri doğrultusunda incelenmiştir.

Nihai uygulama sonrası elde edilen verilerin analizinde pilot uygulama analizlerinde olduğu gibi çoktan seçmeli maddelerin analizinde de TAP programı kullanılmış ve TAP programına ait verilere Ek 5'te yer verilmiştir. Açık uçlu maddelerin analizinde ise Excel'de elde hesaplamalar yapılmıştır. Buna göre açık uçlu maddelerin madde güçlükleri ve madde ayırt edicilikleri her bir puanlayıcı için Eşitlik 9 ve Eşitlik 10'da verilen formüller yardımıyla hesaplanmıştır. Daha sonra elde edilen verilerin ortalamaları alınarak maddelerin güçlükleri ve ayırt edicilikleri belirlenmiştir. Nihai uygulama kapsamında elde edilen verilere ait madde istatistikleri tablo 3.3'te gösterilmiştir.

Tablo 3.3

Nihai Uygulamadaki Testlerin Madde İstatistikleri

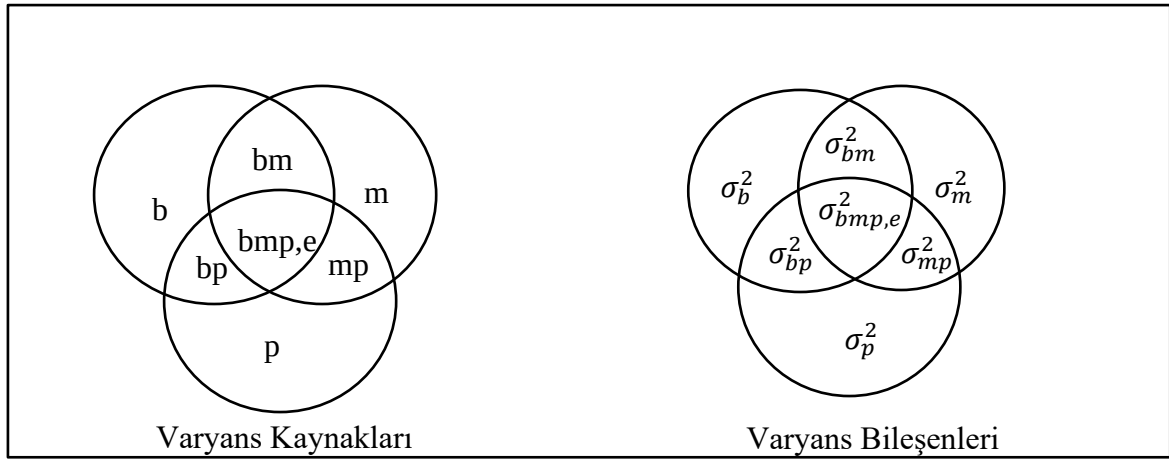
Madde Takımı	Madde No	Çoktan seçmeli maddelere ait madde istatistikleri		Açık uçlu maddelere ait madde istatistikleri	
		p	r	p	r
1	1	0,46	0,57	0,33	0,43
	2	0,43	0,56	0,37	0,60
	3	0,44	0,39	0,20	0,52
	4	0,71	0,51	0,39	0,77
2	5	0,59	0,55	0,27	0,62
	6	0,32	0,53	0,04	0,11
	7	0,29	0,50	0,25	0,36
3	8	0,36	0,49	0,10	0,20
	9	0,37	0,60	0,05	0,14
	10	0,38	0,63	0,30	0,54
4	11	0,51	0,57	0,12	0,38
	12	0,45	0,45	0,14	0,42

Tablo 3.3'teki veriler incelendiğinde çoktan seçmeli maddelerin madde güçlük indekslerinin (p) 0,29 ile 0,71 arasında değer aldığı görülmektedir. Madde güçlük indekslerinin ortalaması alındığında (0,44) maddelerin orta düzeye yakın oldukları söylenebilir. Açık uçlu maddelerin madde güçlüğüne bakıldığında ise değerlerin 0,04 ile 0,39 arasında değiştiği görülmektedir. Açık uçlu maddelerin madde güçlük indekslerinin ortalamasına (0,21) bakıldığında; açık uçlu maddelerin çoktan seçmeli maddelere göre, öğrencilere çok daha zor

geldiği şeklinde yorum yapılabilir. Madde ayırt edicilik indeksleri (r) incelendiğinde çoktan seçmeli maddelerin tümünün ayırt edicilik düzeyleri çok iyi seviyede iken yine açık uçlu maddelerden bazılarının ayırt edicilik düzeylerinin düşük olduğu gözlenmiştir. 6, 8 ve 9. açık uçlu maddeler incelendiğinde bu maddelerin çok zor ve aynı zamanda ayırt ediciliklerinin de düşük olduğu belirtilebilir.

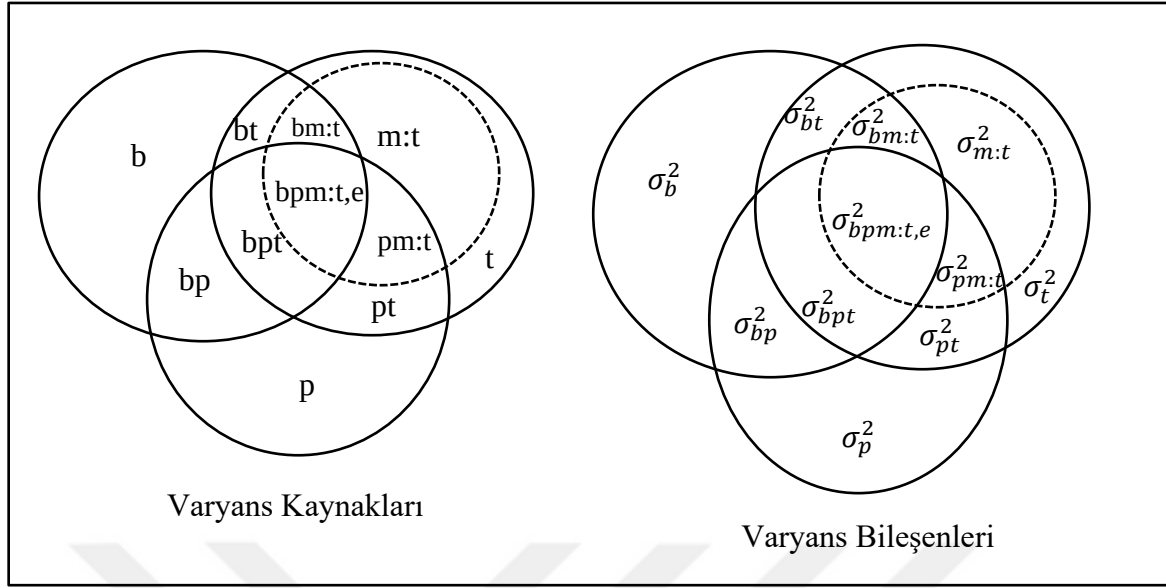
Alt problemler çerçevesinde açık uçlu ve çoktan seçmeli madde takımlarından oluşan testlerde G kuramı çerçevesinde G ve Phi katsayıları hesaplanmıştır. Ayrıca, her bir desen kapsamında madde, madde takımı ve puanlayıcı sayılarından uygun olan değişkenler seçilerek K çalışması yürütülmüştür. Verilerin analizinde EduG programından yararlanılmıştır. G kuramı kapsamında yürütülen G ve K çalışmalarına ait EduG analiz çıktılarına Ek 6'da yer verilmiştir. Alt problemler çerçevesinde gerçekleştirilen işlemler sırasıyla aşağıda yer almaktadır.

1. Açık uçlu maddelerden oluşan madde takımlarının yer aldığı başarı testlerinde sırasıyla aşağıdaki adımlar uygulanmıştır.
 - a. Madde takımı etkisi ele alınmadığında birey (b), madde (m) ve puanlayıcıların (p) birbiriyle çaprazlandığı $b \times m \times p$ tümüyle çaprazlanmış deseni kullanılmıştır. Bu desene ait varyans kaynakları ve varyans bileşenlerinin yer aldığı venn diyagramı aşağıda Şekil 3.1'de gösterilmiştir.



Şekil 3.1. İki yüzeyli çaprazlanmış bxm xp deseni için varyans kaynakları ve varyans bileşenleri. Atılğan, H. (2019). *Genellenebilirlik kuramı ve uygulaması*. Ankara: Anı kaynağından uyarlanmıştır.

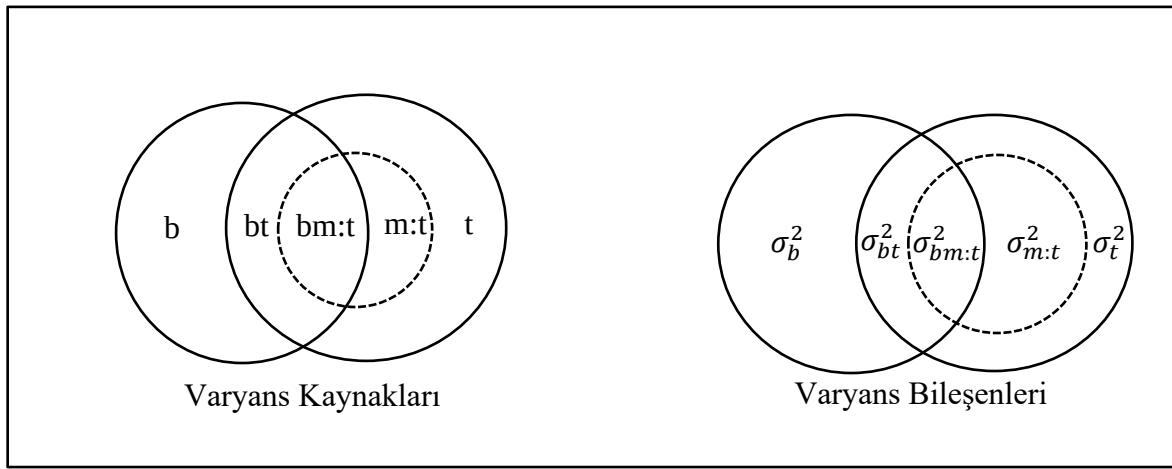
- b. Madde takımı etkisi ele alınmadan oluşturulan bxm xp deseni için; madde sayılarının (6, 9, 15, 18 ve 21 madde) ve puanlayıcı sayılarının (2, 4, 5, 6 ve 7 puanlayıcı) değişiminin G ve Phi katsayılarına etkisini belirlemek için K çalışması yürütülmüştür.
- c. Madde takımı etkisi dikkate alınarak oluşturulan; maddelerin (m) madde takımlarında (t) yuvalandığı, bireylerin (b) ve puanlayıcıların (p) ise bunlarla çaprazlandığı bx(m:t)xp deseni kullanılmıştır. Bu desene ait varyans kaynakları ve varyans bileşenlerine aşağıda Şekil 3.2’de yer verilmiştir.



Şekil 3.2. Üç yüzeyli $bx(m:t)xp$ deseni için varyans kaynakları ve varyans bileşenleri. Atılğan, H. (2019). *Genellenebilirlik kuramı ve uygulaması*. Ankara: Anı kaynağından uyarlanmıştır.

- d. Madde takımı etkisi dikkate alınarak oluşturulan $bx(m:t)xp$ deseni için; madde takımı sayılarının (3, 5, 6, 7 ve 8 madde takımı), madde takımı içindeki madde sayılarının (1, 2, 4, 5 ve 6 madde) ve puanlayıcı sayılarının (2, 4, 5, 6 ve 7 puanlayıcı) değişiminin G ve Phi katsayılarına etkisini belirlemek için K çalışması yürütülmüştür.
2. Çoktan seçmeli maddelerden oluşan madde takımlarının yer aldığı başarı testlerinde sırasıyla aşağıdaki işlemler yapılmıştır.
 - a. Madde takımı etkisi ele alınmadığında; bireylerin (b) maddelerle (m) çaprazlandığı bxm deseni kullanılmıştır. Bu desene ait değişkenlik kaynakları ve varyans bileşenleri Şekil 1.2’de venn diyagramı ile sunulmuştur. Bireyin gözlenen puanlarına ilişkin varyans bileşeninin; bireye ait varyans bileşeni, madde varyans bileşeni ile kesişimlerinden elde edilen birey-madde ve kalan varyans bileşenlerinin toplamından oluşmaktadır (Eşitlik 5).

- b. Madde takımı etkisinin ele alınmadığı bxm deseni için; madde sayılarının değişiminin (6, 9, 15, 18 ve 21 madde) G ve Phi katsayılarına etkisini belirlemek için K çalışması yürütülmüştür.
- c. Madde takımı etkisi dikkate alınarak oluşturulan; maddelerin (m) madde takımlarında (t) yuvalandığı, bireylerin (b) ise bunlarla çaprazlandığı bx(m:t) kısmi yuvalanmış deseni kullanılmıştır. Bu desene ait varyans kaynakları ve varyans bileşenlerine aşağıda Şekil 3.3'te yer verilmiştir.



Şekil 3.3. İki yüzeyli yuvalanmış bx(m:t) deseni için varyans kaynakları ve varyans bileşenleri. Atılgan, H. (2019). *Genellenebilirlik kuramı ve uygulaması*. Ankara: Anı kaynağından uyarlanmıştır.

- d. Madde takımı etkisi dikkate alınarak oluşturulan bx(m:t) deseni için; madde takımı içindeki madde sayılarının (2, 4, 5, 6 ve 7 madde) ve madde takımı sayılarının (2, 3, 5, 6 ve 7 madde takımı) değişiminin G ve Phi katsayılarına etkisini belirlemek için K çalışması yürütülmüştür.
3. Bu alt problemde açık uçlu ve çoktan seçmeli maddelerden oluşan madde takımlarında madde takımı etkisinin ele alınmadığı ve dikkate alındığı durumlarda elde edilen G ve Phi katsayılarına ilişkin tablo oluşturulmuştur.

BÖLÜM IV

BULGULAR VE YORUMLAR

Bu bölümde araştırma sonucunda elde edilen bulgulara ve bu bulgulara ait yorumlara yer verilmiştir.

Birinci Alt Probleme İlişkin Bulgular ve Yorumlar

- a) “Açık uçlu maddelerden oluşan madde takımlarının yer aldığı başarı testlerinde, madde takımı etkisi ele alınmadığında; birey (b), madde (m) ve puanlayıcıların (p) birbiriyle çaprazlandığı bmxmp tümüyle çaprazlanmış desenine ait varyans bileşenleri ile G ve Phi katsayıları nedir?”

Çalışmada 8.sınıfta öğrenim gören 128 öğrencinin 12 tane açık uçlu maddeden oluşan başarı testine verdikleri yanıtlar madde takımı etkisi dikkate alınmadan G kuramı çerçevesinde incelenmiştir. Bu bağlamda birey (b), madde (m) ve puanlayıcıların (p) birbirleriyle çaprazlandığı bmxmp desenine ait G çalışması sonuçları Tablo 4.1’de verilmiştir.

Tablo 4.1

*Açık Uçlu Maddelerden Oluşan Başarı Testinde Madde Takımı Etkisinin Ele Alınmadığı
bxm xp Desenine Ait G Çalışması Sonuçları*

Varyans Kaynağı	Kareler Toplamı	Serbestlik Derecesi	Kareler Ortalaması	Varyans Değeri	Varyans Yüzdesi
b	1352,284	127	10,648	0,238	19,3
p	7,461	2	3,731	0,000	0,0
m	551,671	11	50,152	0,115	9,3
bp	91,816	254	0,361	0,007	0,6
bm	2773,607	1397	1,985	0,571	46,2
pm	92,247	22	4,193	0,031	2,5
bpm,e	763,809	2794	0,273	0,273	22,1
Toplam	5632,895	4607			100%
G Katsayısı	0,81				
Phi Katsayısı	0,78				

Tablo 4.1 incelendiğinde;

- Bireylere (b) ait varyans değerinin 0,238 ile toplam varyansın %19,3'ünü açıklamaktadır. Bu değer en büyük üçüncü varyans değeri olup bireylere ait olan bu varyans bileşeni, bireylerin birbirinden ne derece farklılık gösterdiğini ifade etmektedir. Bu varyans bileşeninin değerinin oldukça yüksek olması beklenir. Değişkenliğin büyük bir kısmının diğer değişkenlik kaynaklarınca açıklandığı görülmektedir.
- Puanlayıcı (p) varyans bileşeni için hesaplanan değerin 0,000 olduğu görülmektedir. Bu değer istenen bir değerdir ve puanlayıcıların puanlamaları arasındaki tutarlılığın mükemmel olduğunu gösterir. Ayrıca burada hazırlanan DPA'nın da iyi hazırlanmış olduğu, böylece puanlayıcıların yaptıkları puanlamalarda çelişkili durumlar ortaya çıkmadığı söylenebilir.
- Madde (m) varyans bileşenine ait hesaplanan değer 0,115 ile toplam varyansın %9,3'ünü açıklamaktadır. Bu durum maddelerin güçlük düzeylerinin farklılık gösterdiğine işaret etmektedir.

- Birey x puanlayıcı (bp) etkileşimi varyans bileşeni için elde edilen değer ise 0,007 ile toplam varyansın %0,6'sını açıklamaktadır. Bu değer, puanlayıcıların bireylere verdikleri puanların puanlayıcılar arasında çok farklılık göstermediği anlamına gelmektedir. Bir başka deyişle, puanlayıcılar bireylere benzer puanları vermişlerdir ki bu da istenen bir durumdur.
 - Birey x madde (bm) etkileşimi varyans bileşeni incelendiğinde hesaplanan değer 0,571 ile en büyük varyans değeridir ve toplam varyansın %46,2'sini açıklamaktadır. Bu değer, maddelerin güçlük düzeylerinin bireyden bireye farklılık gösterdiğini ifade etmektedir.
 - Puanlayıcı x madde (pm) etkileşimi varyans bileşeni için hesaplanan değer 0,031 ile toplam varyansın %2,5'ini açıklamaktadır. Bu değer puanlayıcıların maddelere verdikleri puanların değişkenliğini belirtmektedir. Bu değer düşük olması istenen bir durumdur.
 - En yüksek ikinci varyans değeri 0,273 ile artık bileşene aittir. Toplam varyansın %22,1'ini açıklamaktadır. Bu değer bireyler, maddeler ve puanlayıcılar arası etkileşim ile bu çalışmada sistematik ya da sistematik olmayan hata kaynakları olduğunun göstergesidir.
 - Son olarak, G kuramı çalışması sonucunda elde edilen analizlerde G katsayısının 0,81 ve Phi katsayısının 0,78 olarak hesaplandığı görülmektedir. Elde edilen bu değerler 0,70'in oldukça üzerinde olduğundan kabul edilebilir değerlerdir.
- b) “Açık uçlu maddelerden oluşan madde takımlarının yer aldığı başarı testlerinde, madde takımı etkisi ele alınmadığında; birey (b), madde (m) ve puanlayıcıların (p) birbiriyle çaprazlandığı bmxp deseninde, madde ve puanlayıcı sayısının artırılıp azaltılmasına yönelik yapılan karar çalışmasında G ve Phi katsayıları nasıldır?”**

Açık uçlu maddelerden oluşan madde takımlarının yer aldığı başarı testlerine ait analizler G kuramı kapsamında yapılan K çalışmaları ile yürütülmüştür. Madde takımı etkisinin ele

alınmadığı bireylerin (b) maddeler (m) ve puanlayıcılar (p) ile tümüyle çaprazlandığı bmxpxp deseninde madde ve puanlayıcı sayısının değiştirilmesi elde edilen G ve Phi katsayılarına Tablo 4.2’de yer verilmiştir.

Tablo 4.2

Açık Uçlu Maddelerden Oluşan Başarı Testinde Madde Takımı Etkisinin Ele Alınmadığı bmxpxp Desenine Ait K Çalışması Sonuçları

	Madde Sayısı	G Katsayısı	Phi Katsayısı
Birey sayısı: 128	6	0,68	0,64
	9	0,76	0,73
	12*	0,81*	0,78*
Puanlayıcı sayısı: 3	15	0,84	0,81
	18	0,86	0,84
	21	0,88	0,86
	Puanlayıcı Sayısı		
Birey sayısı: 128	2	0,79	0,76
	3*	0,81*	0,78*
	4	0,81	0,78
Madde sayısı: 12	5	0,82	0,79
	6	0,82	0,79
	7	0,82	0,79

* Mevcut çalışmanın durum çalışmasına ait verilerini ifade etmektedir.

Tablo 4.2’de verilen açık uçlu maddelerden oluşan madde takımlarının yer aldığı başarı testinin bmxpxp desenine ait K çalışmaları incelendiğinde; birey ve puanlayıcı sayısı sabit tutulup, madde sayısı artırıldığında G ve Phi katsayılarının da arttığı gözlenmektedir. Birey ve madde sayısı sabit tutulup, puanlayıcı sayısı artırıldığında ise yine güvenilirlik katsayılarının arttığı fakat madde sayısındaki artış kadar etkilenmediği görülmektedir. Puanlayıcı sayısındaki artış incelendiğinde; üç puanlayıcının üzerine çıkıldığında artış miktarında çok az bir farklılık olduğu göze çarpmaktadır. Hatta iki puanlayıcıyla bile G katsayısının değeri 0,79 olarak hesaplanmıştır. Puanlayıcı sayısındaki artışın güvenilirlik katsayıları üzerinde çok büyük bir değişikliğe etkisi olmadığı görülmektedir. Bu durumda zaman, ekonomiklik ve pratiklik de göz önüne alınırsa bu araştırma kapsamında iki puanlayıcının yeterli olabileceği söylenebilir.

- c) “Açık uçlu maddelerden oluşan madde takımlarının yer aldığı başarı testlerinde, madde takımı etkisi dikkate alınarak oluşturulan; maddelerin (m) madde takımlarında (t) yuvalandığı, bireylerin (b) ve puanlayıcıların (p) ise bunlarla çaprazlandığı $bx(m:t)xp$ deseninin varyans bileşenleri ile G ve Phi katsayıları nedir?”

Madde takımı etkisinin dikkate alındığı $bx(m:t)xp$ desenine ait G çalışması sonucu kestirilen varyans bileşenleri ve toplam varyansı açıklama oranları tablo 4.3’te gösterilmiştir.

Tablo 4.3

Açık Uçlu Maddelerden Oluşan Başarı Testinde Madde Takımı Etkisinin Dikkate Alındığı $bx(m:t)xp$ Desenine Ait G Çalışması Sonuçları

Varyans Kaynağı	Kareler Toplamı	Serbestlik Derecesi	Kareler Ortalaması	Varyans Değeri	Varyans Yüzdesi
b	1352,284	127	10,648	0,199	16,0
p	7,461	2	3,731	-0,001	0,0
t	145,362	3	48,454	-0,005	0,0
m:t	406,309	8	50,789	0,119	9,6
bp	91,816	254	0,361	0,003	0,3
bt	1318,138	381	3,460	0,218	17,6
bm:t	1455,469	1016	1,433	0,392	31,6
pt	30,228	6	5,038	0,003	0,2
pm:t	62,019	16	3,876	0,028	2,3
bpt	243,605	762	0,320	0,021	1,7
bpt:m,e	520,203	2032	0,256	0,256	20,6
Toplam	5632,895	4607			100%
G Katsayısı	0,67				
Phi Katsayısı	0,65				

Tablo 4.3 incelendiğinde;

- Bireylere ait varyans bileşeninin değeri 0,199 ile toplam varyansın %16’sını açıklamaktadır. Bireylerin birbirinden sistematik şekilde farklılık gösterdiğinin bir göstergesi olan bu değerin yüksek olması istenen bir durumdur.
- Puanlayıcı ve madde takımı varyans bileşenlerine ait hesaplanan değerler negatif (sırasıyla -0,001 ve -0,005) çıkarak toplam varyansın açıklanmasına katkıda

bulunmamışlardır. Bu değerlerin sıfıra çok yakın olması puanlayıcıların yaptıkları puanlamaların ve madde takımlarının birinden diğerine madde güçlüğüünün farklılık göstermediğini belirtmektedir. Ayrıca bu değerlerin negatif olması küçük bir örneklem üzerinde çalışılmış olunabileceğinin göstergesi olarak kabul edilmektedir (Güler, Kaya Uyanık & Teker, 2012, s.69).

- Maddelerin madde takımları içinde yuvalandığı m:t varyans bileşenine ait hesaplanan değer 0,119 ile toplam varyansın %9,6'sını açıklamaktadır. Bu değer, madde takımları içindeki maddelerin güçlük düzeyleri arasında biraz da olsa farklılık olduğunun göstergesidir.
- Birey x puanlayıcı (bp) ile puanlayıcı x madde takımı (pt) varyans bileşenlerinin 0,003'lük bir değere sahip oldukları görülmektedir. Bu değerler ile toplam varyansın sırasıyla %0,3'ü ile %0,2'sini açıklamaktadırlar. Değerlerin sıfıra çok yakın olması puanlayıcıların aynı bireyler ve aynı madde takımları üzerindeki puanlamalarında farklılık olmadığını göstermektedir. Bu durum puanların objektifliği açısından istenen bir durumdur.
- Birey x madde takımına ait varyans bileşeni incelendiğinde ise 0,218 değeri ile toplam varyansın %17,6'sını açıklamakta ve bu değer en büyük üçüncü yüzde değeri olduğu görülmektedir. Buradan birey madde takımı etkileşiminden kaynaklı farklılıkların yer aldığı belirtilebilir. Yani bireylerin madde takımlarından elde ettikleri puanların farklılaştığı ifade edilebilir.
- Birey x madde ortak etkileşimin madde takımı içindeki bm:t varyans bileşenine ait hesaplanan değer 0,392 ile en büyük değere sahiptir. Bu değer toplam varyansın tek başına %31,6'sını açıklamaktadır. Bu değer birey madde ortak etkileşimi madde takımı içinde ele alındığında farklılık göstermekte olduğunu belirtmektedir.
- Artık bileşene ait varyans değeri incelendiğinde ise 0,256 ile toplam varyansın %20,6'sını açıkladığı görülmektedir. Bu değer en büyük ikinci varyans değeri olup birey, puanlayıcı ve madde takımları içinde maddelerin etkileşimi ile birlikte

sistematiik ya da sistematiik olmayan farklı hata kaynaklarının olduėunun bir g stergesidir.

- Son olarak G kuramı analizleri ile elde edilen $bx(m:t)xp$ desenine ait G katsayısı 0,67 ve Phi katsayısı 0,65 olarak elde edilmiřtir. G ve Phi katsayıları madde takımı etkisinin ele alınmadığı duruma kıyasla daha d ř k elde edilmiřtir.
- d) “A ık u lu maddelerden oluřan madde takımlarının yer aldıėı bařarı testlerinde, madde takımı etkisinin dikkate alındığı; maddelerin (m) madde takımlarında (t) yuvalandıėı, bireylerin (b) ve puanlayıcıların (p) ise bunlarla  aprazlandıėı $bx(m:t)xp$ deseninde, madde takımı, madde takımı i indeki madde sayısı ve puanlayıcı sayısının artırılıp azaltılmasına y nelik yapılan karar  alıřmasında G ve Phi katsayıları nasıldır?”

Son olarak a ık u lu maddelerden oluřan madde takımlarının yer aldıėı bařarı testlerinin madde takımı etkisinin ele alındığı $bx(m:t)xp$ deseninde madde takımı, madde sayısı ve puanlayıcı sayısının deėiřtirilmesine y nelik y r t len K  alıřması sonu larına Tablo 4.4’te yer verilmiřtir.

Tablo 4.4

*Açık Uçlu Maddelerden Oluşan Başarı Testinde Madde Takımı Etkisinin Dikkate Alındığı
bx(m:t)xp Desenine Ait K Çalışması Sonuçları*

	Madde Takımı Sayısı	G Katsayısı	Phi Katsayısı
	3	0,61	0,58
Birey sayısı: 128	4*	0,67*	0,65*
Puanlayıcı sayısı: 3	5	0,72	0,70
Madde takımı içindeki	6	0,75	0,73
madde sayısı: 3	7	0,78	0,76
	8	0,80	0,78
	Madde Takımı İçindeki Madde Sayısı		
	1	0,53	0,49
	2	0,63	0,60
Birey sayısı: 128	3*	0,67*	0,65*
Puanlayıcı sayısı: 3	4	0,69	0,67
Madde takımı sayısı: 4	5	0,71	0,69
	6	0,72	0,71
	Puanlayıcı Sayısı		
	2	0,66	0,64
Birey sayısı: 128	3*	0,67*	0,65*
Madde takımı sayısı: 4	4	0,68	0,65
Madde takımı içindeki	5	0,68	0,66
madde sayısı: 3	6	0,68	0,66
	7	0,68	0,66

* Mevcut çalışmanın durum çalışmasına ait verilerini ifade etmektedir.

Tablo 4.4 incelendiğinde; diğerleri sabit tutularak sırasıyla madde takımı, madde takımı içindeki madde sayısı ve puanlayıcı sayısı artırıldığında G ve Phi katsayılarının da arttığı görülmektedir. Açık uçlu maddelerden oluşan madde takımlarında güvenilirlik katsayısına madde takımı sayısındaki artışın madde takımı içindeki madde sayısı artışından daha fazla etki ettiği gözlenmiştir. G katsayısının; madde takımı içindeki madde sayısı artırıldığında 5'er maddeden oluşan 4 madde takımı (toplam 20 madde) ile 0,70'in üzerine çıktığı gözlemlenirken, madde takımı sayısı artırıldığında 3'er maddeden oluşan 5 madde takımı (toplam 15 madde) ile 0,70'in üzerine çıktığı görülmektedir. Puanlayıcı sayısındaki artışın

ise, güvenilirlik katsayılarını çok büyük oranda etkilemediği göze çarpmaktadır. Bu durumda puanlayıcı bulmanın da getirdiği güçlükler de göz önüne alınırsa, pratiklik açısından iki puanlayıcının bile yeterli olacağı söylenebilir.

İkinci Alt Probleme İlişkin Bulgular ve Yorumlar

- a) “Çoktan seçmeli maddelerden oluşan madde takımlarının yer aldığı başarı testlerinde, madde takımı etkisi ele alınmadığında; bireylerin (b) maddelerle (m) çaprazlandığı bxm deseninin varyans bileşenleri ile G ve Phi katsayıları nedir?”

Çalışma kapsamında 128 8.sınıf öğrencisinin çoktan seçmeli 12 maddeden oluşan başarı testine verdikleri yanıtların madde takımı etkisi dikkate alınmadığı bxm desenine ait G çalışması sonucu kestirilen varyans bileşenleri ve toplam varyans açıklama oranları Tablo 4.5’te gösterilmiştir.

Tablo 4.5

Çoktan Seçmeli Maddelerden Oluşan Başarı Testinde Madde Takımı Etkisinin Ele Alınmadığı bxm Desenine Ait G Çalışması Sonuçları

Varyans Kaynağı	Kareler Toplamı	Serbestlik Derecesi	Kareler Ortalaması	Varyans Değeri	Varyans Yüzdesi
b	100,093	127	0,788	0,050	20,2
m	19,898	11	1,809	0,013	5,1
bm,e	258,852	1397	0,185	0.185	74,7
Toplam	378,843	1535			100%
G Katsayısı	0,76				
Phi Katsayısı	0,75				

Tablo 4.5 incelendiğinde;

- Bireyler için kestirilen varyans bileşeninin 0,050 değeri ile toplam varyansın %20,2’sini oluşturduğu görülmektedir. Bu değer de en büyük ikinci yüzdeliğe sahiptir. Bireylere ait olan bu varyans bileşeninin oldukça yüksek olması bireyler

arasında sistematik olarak farklılık olduğunun göstergesidir ve bu durum istenen bir durumdur.

- Madde varyans bileşeni için hesaplanan değer 0,013 değeri ile toplam varyansın %5,1'ini açıklamaktadır. Bu değer ne kadar yüksek olursa maddelerin güçlüklerinin de bir o kadar farklı olduğu anlamına gelir. Burada madde güçlüklerinin çok farklılaşmadığını söylemek mümkündür.
 - Son olarak en yüksek varyans değerinin 0,185 ile artık bileşene ait olduğu görülmektedir. Bu değer toplam varyansın %74,7'sini oluşturmaktadır. Bu durum birey madde etkileşimi ile bu çalışmada ölçülemeyen sistematik ya da sistematik olmayan hata kaynaklarının kontrol edilmediğinin bir göstergesidir. Artık bileşene ait olan bu varyansın oldukça düşük olması beklenir.
 - Çalışmada ele alınan 128 bireye uygulanmış 12 maddeden oluşan testin G kuramı analizleri sonucunda G katsayısı 0,76 ve Phi katsayısı 0,75 olarak elde edilmiştir. Bu durumda elde edilen güvenilirlik katsayılarının yeterli düzeyde olduğu söylenebilir.
- b) “Çoktan seçmeli maddelerden oluşan madde takımlarının yer aldığı başarı testlerinde, madde takımı etkisinin ele alınmadığı, bireylerin (b) maddelerle (m) çaprazlandığı bxm deseninde; madde sayısının artırılıp azaltılmasına yönelik yapılan karar çalışmasında G ve Phi katsayıları nasıldır?”**

Çoktan seçmeli maddelerden oluşan madde takımlarının yer aldığı başarı testlerine ait madde takımı etkisinin ele alınmadığı bxm desenine ait madde sayısının değiştirilmesine yönelik yürütülen K çalışması sonuçlarına Tablo 4.6'da yer verilmiştir.

Tablo 4.6

Çoktan Seçmeli Maddelerden Oluşan Başarı Testinde Madde Takımı Etkisinin Ele Alınmadığı bxm Desenine Ait K Çalışması Sonuçları

	Madde Sayısı	G Katsayısı	Phi Katsayısı
Birey Sayısı: 128	6	0,62	0,60
	9	0,71	0,70
	12*	0,76*	0,75*
	15	0,80	0,79
	18	0,83	0,82
	21	0,85	0,84

* Mevcut çalışmanın durum çalışmasına ait verilerini ifade etmektedir.

Tablo 4.6 incelendiğinde çoktan seçmeli maddelerden oluşan başarı testlerinde; bireylerin (b) maddelerle (m) çaprazlandığı bxm desenine ait K çalışması sonucu elde edilen G ve Phi katsayılarına bakıldığında, madde sayısı attıkça elde edilen güvenilirlik katsayılarının da arttığı gözlenmiştir. Madde sayısında 6 maddeden 15 maddeye çıkıldığında G katsayısının da 0,62'den 0,80'e çıktığı görülmektedir. Bu durumda testin amacına göre karar vermek daha doğru olacaktır. Eğer yüksek riskli bir sınav ise; güvenirliliğin 0,80 ve üzeri olması önerilmektedir (Nunnally, 1967; akt. Henson, 2001). Bu durumda daha yüksek güvenilirlik için 15 madde ve üzeri tercih edilmeli iken, daha düşük riskli sınavlarda ise 0,70 ve üzeri güvenilirlik değeri yeterli olabileceğinden 9 ile 15 madde arası tercihte bulunmak yeterli olacaktır.

- c) “Çoktan seçmeli maddelerden oluşan madde takımlarının yer aldığı başarı testlerinde, madde takımı etkisi dikkate alınarak oluşturulan; maddelerin (m) madde takımlarında (t) yuvalandığı, bireylerin (b) ise bunlarla çaprazlandığı bx(m:t) kısmi yuvalanmış deseninin varyans bileşenleri ile G ve Phi katsayıları nedir?”

Çoktan seçmeli maddelerden oluşan madde takımlarının madde takımı etkisi dikkate alınarak oluşturulan, maddelerin madde takımlarında yuvalandığı bireylerin ise bunlarla

çaprazlandığı $bx(m:t)$ desenine ait G çalışması sonucunda kestirilen varyans bileşenleri ve toplam varyansı açıklama oranları Tablo 4.7’de gösterilmiştir.

Tablo 4.7

Çoktan Seçmeli Maddelerden Oluşan Başarı Testinde Madde Takımı Etkisinin Dikkate Alındığı $bx(m:t)$ Desenine Ait G Çalışması Sonuçları

Varyans Kaynağı	Kareler Toplamı	Serbestlik Derecesi	Kareler Ortalaması	Varyans Değeri	Varyans Yüzdesi
b	100,093	127	0,788	0,047	18,8
t	7,929	3	2,643	0,003	1,1
m:t	11,969	8	1,496	0,010	4,2
bt	86,154	381	0,226	0,019	7,5
bm:t,e	172,698	1016	0,170	0,170	68,3
Toplam	378,843	1535			100%
G Katsayısı	0,71				
Phi Katsayısı	0,70				

Tablo 4.7 incelendiğinde;

- Birey (b) ana etkisine ait varyans değerinin 0,047 ile toplan varyansın %18,8’ini açıkladığı görülmektedir. Bu değer tabloda okunan en büyük ikinci varyans yüzdesi olarak karşımıza çıkmaktadır ve bu istenen bir durumdur. Bireyler arasında sistematik olarak farklılık olduğunun göstergesidir.
- Madde takımı (t) ana etkisi için kestirilen varyans değeri 0,003 ile toplan varyansın %1,1’ini açıklamaktadır. Gözlenen bu değer düşük olması madde takımlarının güçlük düzeylerinin bir madde takımından diğerine çok fazla farklılık göstermediği anlamına gelmektedir.
- Maddelerin madde takımları içinde yuvalandığı m:t etkisine ait varyans bileşeni 0,010 değeri ile toplam varyansın %4,2’sini açıklamaktadır. Bu değer sıfıra yaklaştıkça madde takımları içindeki maddelerin güçlük düzeylerinin birbirinden farklı olmadığı şeklinde yorumlanabilir. Çalışmadan elde edilen değer de düşük

olması nedeniyle aynı madde takımları içindeki maddelerin güçlük düzeylerinin birbirine yakın olduğu söylenebilir.

- Birey x madde takımı (bt) etkisine bakıldığında kestirilen varyans bileşenin 0,019 değeri ile toplam varyansın %7,5'ini açıkladığı görülmektedir. Kestirilen bu varyans değeri üçüncü en büyük değere sahip olup birey madde takımı etkileşiminden kaynaklı farklılıklar olduğu anlamına gelmektedir.
 - Son olarak, en büyük varyans değeri 0,170 ile birey-madde-madde takımı etkisi olan artık bileşene (bm:t,e) aittir. Bu değer toplam varyansın %68,3'ünü açıklamaktadır. Bu değer; bireyler, maddeler ve madde takımları arası etkileşim ile bu çalışmada ölçülemeyen sistematik ya da sistematik olmayan değişkenlik kaynaklarının oldukça büyük olduğunun göstergesidir. Yine de madde takımı etkisi ele alınmadığında elde edilen artık bileşenin toplam varyansı açıklama yüzdesi değerine (%74,7) göre daha düşük bir değer olduğunu görmekteyiz. Bu da bize madde takımı etkisi ele alındığında daha fazla varyans kaynağı olduğundan toplam varyansı açıklama yüzdesinin bu varyans kaynaklarına bölündüğünü ve artık bileşene ait varyans yüzdesinin düştüğünü göstermektedir.
 - Güvenirlik katsayıları incelendiğinde ise, G kuramı analizleri sonucunda elde edilen G katsayısının 0,71 ve Phi katsayısının 0,70 değeri ile birbirlerine oldukça yakın değerler olduğu görülmektedir. Ayrıca bu değerlerin literatürde kabul edilebilir değerler olduğunu söylemek mümkündür.
- d)** “Çoktan seçmeli maddelerden oluşan madde takımlarının yer aldığı başarı testlerinde, madde takımı etkisinin dikkate alındığı, maddelerin (m) madde takımlarında (t) yuvalandığı, bireylerin (b) ise bunlarla çaprazlandığı $bx(m:t)$ kısmi yuvalanmış deseninde; madde takımı sayılarının ve madde takımı içindeki madde sayılarının artırılıp azaltılması ile yapılan karar çalışmalarında G ve Phi katsayıları nasıldır?”

Çoktan seçmeli maddelerden oluşan başarı testlerinde madde takımı etkisinin ele alındığı, maddelerin (m) madde takımları (t) içinde yuvalandığı ve bunların da bireyler (b) ile çaprazlandığı $bx(m:t)$ deseninde madde takımı ve madde sayılarının değiştirilmesine yönelik yürütülen K çalışmasına ait sonuçlar Tablo 4.8’de gösterilmiştir.

Tablo 4.8

Çoktan Seçmeli Maddelerden Oluşan Başarı Testinde Madde Takımı Etkisinin Dikkate Alındığı $bx(m:t)$ Desenine Ait K Çalışması Sonuçları

	Madde Takımı Sayısı	G Katsayısı	Phi Katsayısı
Birey sayısı: 128	2	0,55	0,53
	3	0,65	0,63
	4*	0,71*	0,70*
Madde takımı içindeki madde sayısı: 3	5	0,76	0,74
	6	0,79	0,77
	7	0,81	0,80
	Madde Takımı İçindeki Madde Sayısı		
Birey sayısı: 128	2	0,64	0,63
	3*	0,71*	0,70*
	4	0,75	0,74
Madde takımı sayısı: 4	5	0,78	0,76
	6	0,80	0,78
	7	0,81	0,80

* Mevcut çalışmanın durum çalışmasına ait verilerini ifade etmektedir.

Tablo 4.8’de çoktan seçmeli maddelerden oluşan başarı testinin $bx(m:t)$ desenine ait K çalışması incelendiğinde; birey sayısı ve madde takımı içindeki madde sayısı sabit tutulup, madde takımı sayısı artırıldığında G ve Phi katsayılarının da arttığı görülmüştür. Diğer yandan birey ve madde takımı sayısı sabit tutulup, madde takımı içindeki madde sayısı artırıldığında yine güvenilirlik katsayılarının arttığı gözlenmektedir. Fakat madde takımı sayısındaki artışın madde takımı içindeki madde sayısındaki artışa göre güvenilirlik katsayılarını daha çok etkilediği gözlenmektedir. Bu sonuç açık uçlu maddelerden oluşan madde takımları ile elde edilen sonuçlar ile benzerdir. Öyle ki; 3’er maddeden oluşan 7

madde takımının (toplam 21 madde) G katsayısı 0,81 iken, 7’şer maddeden oluşan 4 madde takımının (toplam 28 madde) G katsayısı da 0,81 olarak hesaplanmıştır. Daha çok madde ile aynı G katsayısının elde edilmesi madde takımı artışının daha etkili olduğunun göstergesidir.

Üçüncü Alt Probleme İlişkin Bulgular ve Yorumlar

“Açık uçlu ve çoktan seçmeli maddelerden oluşan madde takımlarının yer aldığı başarı testlerinde madde takımı etkisinin ele alınmadığı ve dikkate alındığı durumlarda G ve Phi katsayıları nasıl değişmektedir?”

Çoktan seçmeli ve açık uçlu maddelerden oluşan madde takımlarının yer aldığı başarı testlerine ait G kuramı kapsamında elde edilen güvenirlik katsayıları Tablo 4.9’da verilmiştir.

Tablo 4.9

Açık Uçlu ve Çoktan Seçmeli Maddelerden Oluşan Başarı Testlerinde Madde Takımı Etkisine Dayalı Elde Edilen Güvenirlik Katsayıları

		Açık uçlu maddelerden oluşan başarı testleri		Çoktan seçmeli maddelerden oluşan başarı testleri	
		Madde takımı etkisinin ele alınmadığı durum: bmxp	Madde takımı etkisinin dikkate alındığı durum: bx(m:t)xp	Madde takımı etkisinin ele alınmadığı durum: bxm	Madde takımı etkisinin dikkate alındığı durum: bx(m:t)
Güvenirlik	G	0,81	0,67	0,76	0,71
Katsayıları	Phi	0,78	0,65	0,75	0,70

Tablo 4.9 incelendiğinde üçer maddelik dört madde takımından oluşan benzer içerikte hazırlanmış açık uçlu ve çoktan seçmeli maddelerin yer aldığı başarı testlerine ait madde takımı etkisinin ele alınmadığı ve dikkate alındığı durumlar için G ve Phi katsayıları görülmektedir. Açık uçlu maddelerden oluşan başarı testlerinin madde takımı etkisinin ele alınmadığı bmxp desenine ait G katsayısı 0,81 ve Phi katsayısı 0,78 iken; madde takımı etkisinin dahil edildiği bx(m:t)xp desenine ait G katsayısı 0,67 ve Phi katsayısı 0,65 olarak hesaplanmıştır. G ve Phi katsayılarının birbirine yakın oldukları görülmekle birlikte madde

takımı etkisi ele alınmadığı durumda güvenilirlik katsayılarının daha yüksek kestirildiği gözlenmiştir. Çoktan seçmeli maddelerden oluşan başarı testlerinin madde takımı etkisinin ele alınmadığı bxm desenine ait G katsayısı 0,76 ve Phi katsayısı 0,75 iken; madde takımı etkisinin dahil edildiği bx(m:t) desenine ait G katsayısı 0,71 ve Phi katsayısı 0,70 olarak hesaplanmıştır. Açık uçlu maddelerden oluşan başarı testlerinde görüldüğü gibi çoktan seçmeli maddelerden oluşan başarı testlerinde de madde takımı etkisinin güvenilirlik katsayıları arasında farklılığa neden olduğu belirlenmiştir. Yalnız şuna dikkat edilmelidir ki açık uçlu maddelerden oluşan başarı testinde, madde takımı etkisi ele alınıp/alınmadığı durumda oluşan G katsayısındaki değişimin (0,14); çoktan seçmeli maddelerden oluşan başarı testinde, madde takımı etkisi ele alınıp/alınmadığı durumda elde edilen G katsayısındaki değişime (0,05) göre çok daha fazla olduğu görülmektedir. Buna göre madde takımlarının ele alınıp/alınmadığı durumlarda kestirilen G katsayısındaki değişimin madde türüne göre de büyük farklılık gösterdiği söylenebilir. Madde takımı etkisi ele alındığı zaman güvenilirlik katsayılarının daha düşük kestirildiği görülmekle birlikte G katsayısındaki değişimde madde türünün de etkisinin göz ardı edilemeyeceği görülmektedir.

Sonuç olarak; madde takımı etkisinin dikkate alındığı durumlarda yapılan güvenilirlik kestirimlerinin, madde takımı etkisinin ele alınmadığı durumlarda yapılan güvenilirlik kestirimlerine göre farklılık yarattığı görülmüştür. Yapılan bu çalışmada elde edilen sonuçlar alan yazındaki bazı çalışmaların bulguları ile de uyum göstermektedir (Lee ve Frisbie, 1999; Lee, Brennan ve Frisbie, 2000; Taşdelen Teker, 2014; Kaya Uyanık ve Ertuna, 2022).

BÖLÜM V

SONUÇ, TARTIŞMA VE ÖNERİLER

Bu bölümde araştırma sonucunda ulaşılan bulgulardan elde edilen sonuç, tartışma ve önerilere yer verilmiştir.

Birinci Alt Probleme İlişkin Sonuç ve Tartışma

Açık uçlu maddelerden oluşan madde takımlarının yer aldığı başarı testlerinde madde takımı etkisinin ele alınmadığı ve dikkate alındığı durumlara yönelik oluşturulan desenlerde genellenebilirlik kuramı çerçevesinde G ve Phi katsayıları kestirilmiştir. Açık uçlu maddelerden oluşan başarı testlerine ait bulgular incelendiğinde; çoktan seçmeli maddelerden oluşan madde takımlarının yer aldığı başarı testlerinde olduğu gibi; madde takımı etkisinin ele alınmadığı ve dikkate alındığı durumlarda farklı güvenirlik katsayıları kestirilmiştir. Bu sonuç objektif puanlanabilirliği düşük olan açık uçlu maddelerden oluşan madde takımlarında beklendik bir durum olarak belirtilebilir (Kaya Uyanık & Ertuna, 2022). Madde takımı etkisinin ele alınmadığı durumda madde takımı içinde yer alan maddeler arası korelasyondan dolayı testin güvenirlik değerini olduğundan yüksek kestirmektedir (Lee & Park, 2012).

Yürütülen K çalışmaları sonucunda açık uçlu maddelerden oluşan başarı testlerinde birey ve puanlayıcı sayısı sabit tutularak madde sayısı artırıldığında, elde edilen güvenirlik

katsayılarında da artış gözlenmiştir. Madde sayılarındaki artış aynı zamanda testin güvenilirliğini de artıracığından (Baykul, 2000; Turgut, 1992) bu beklendik bir sonuçtur. Ancak güvenilirlik katsayılarının da belli bir madde sayısına kadar yükselerek arttığı, belli bir madde sayısından sonra ise artışın eskisi kadar katkı sağlamadığı görülmüştür. Bu sebeple öncelikle uygulanacak testin yüksek riskli mi yoksa düşük riskli mi olacağına karar verilmelidir. Ardından testin amacına hizmet edecek seviyedeki güvenilirlik katsayılarına göre madde sayısı tercih edilmesi uygun olacaktır. Yüksek riskli testlerde güvenilirlik değeri daha yüksek kabul edilmektedir (Nunnally, 1967; akt. Henson, 2001).

Elde edilen bulgulardan ulaşacağımız sonuçlardan birisi de açık uçlu maddelerden oluşan madde takımlarında madde takımı etkisinin ele alınması gerektiğidir. Bu nedenle madde takımları içindeki madde sayısının artırılmasından ziyade, madde takımı sayısının artırılması daha etkili olacaktır. Ayrıca madde takımı sayısındaki artış ile madde sayısındaki artışlar kıyaslandığında madde takımındaki artışın testteki madde sayısının artışına daha yüksek katkısı olduğu ifade edilebilir. Birey, madde ve madde takımı sayıları sabit tutularak puanlayıcı sayısındaki artışa baktığımızda güvenilirliğe çok fazla katkısı olduğunu söylemek söz konusu değildir. Şartlar da göz önüne alındığında hem fazla sayıda puanlayıcı bulmanın zor ve zahmetli olmasından hem de zaman ve pratiklik açısından iki puanlayıcının yeterli olacağı sonucuna ulaşılmıştır. Bu sonuç Kaya Uyanık ve Ertuna'nın (2022) sonuçları ile paralellik göstermektedir. Taşdelen Teker, Şahin ve Baytemir (2016) öğrencilerin iletişim becerilerinin 5 li dereceleme ölçeği ile değerlendirildiği çalışmalarında gerçekleştirilen K çalışması sonucunda da iki puanlayıcı ile yeterli güvenilirlik değerinin elde edildiği raporlanmıştır. Bu çalışmada tüm açık uçlu maddeler için ayrıntılı olarak geliştirilen dereceli puanlama anahtarının kullanılması da objektifliği artırabileceğinden (Moskal & Leydens, 2000) puanlayıcıdan kaynaklı hatayı da azaltmış olabilir. Hem madde takımı etkisinin ele alınmadığı durumda hem de dikkate alındığı durumda birey puanlayıcı varyansının çok düşük elde edilmesi de bu sonucu destekler niteliktedir.

Sonuç olarak, açık uçlu maddelerden oluşan madde takımlarının yer aldığı başarı testlerinde madde takımı etkisi dikkate alınıp alınmadığı durumda G çalışması sonucu kestirilen G ve Phi katsayılarının birbirinden oldukça farklı olduğu görülmektedir. Bu sebeple açık uçlu maddelerden oluşan madde takımları ile yapılan çalışmalarda daha doğru güvenilirlik kestirimleri için, madde takımı etkisinin mutlaka göz önünde bulundurulması gerektiği sonucuna ulaşılmıştır.

İkinci Alt Probleme İlişkin Sonuç ve Tartışma

Çoktan seçmeli maddelerden oluşan madde takımlarının yer aldığı başarı testlerinde madde takımı etkisinin ele alınmadığı ve dikkate alındığı durumlara yönelik oluşturulan desenlerde genellenebilirlik kuramı çerçevesinde G ve Phi katsayıları kestirilmiştir. Elde edilen bulgular incelendiğinde; açık uçlu maddelerden oluşan madde takımlarının yer aldığı başarı testlerinde olduğu gibi, madde takımı etkisinin dikkate alınıp alınmama durumlarında farklı G ve Phi katsayıları kestirilmiştir.

Çoktan seçmeli maddelerden oluşan testte de açık uçlu maddelerden oluşan testin sonucuna benzer şekilde madde takımı etkisi ele alınmadığı durumda güvenilirlik katsayıları daha yüksek kestirilirken, madde takımı etkisi ele alındığında elde edilen güvenilirlik katsayıları daha düşük olarak kestirilmiştir. Bu sonuç çoktan seçmeli maddelerden oluşan madde takımı ve güvenilirlik etkisinin incelendiği çalışmaların sonuçları ile (Hendrickson, 2001; Lee & Park, 2012; Sireci, Thissen & Wainer, 1991; Taşdelen Teker, 2014; Thissen, Steinberg & Mooney, 1989; Wainer, 1995) uyumlu olarak elde edilmiştir. Madde takımı etkisinin ele alınmadığı durumlar sonuçlarda yanlılığa ve güvenilirlik değerinin daha yüksek kestirimine neden olabilmektedir. Madde takımı etkisinin ele alınmadığı durumda maddeler arası ilişkilerin yüksek olması güvenilirliğin olması gerektiğinden daha yüksek kestirilmesine neden olabilir. Dolayısıyla aynı madde takımı içinde yer alan maddeler arası korelasyonun yüksek olması da homojenliğe katkı sağlayacaktır. Madde takımı etkisinin dikkate alınması durumunda ise madde takımlarının içeriklerinin farklı olması heterojenliği sağlayacaktır. Bu

da madde takımı etkisi ele alındığında güvenilirliğin ele alınmadığı duruma göre düşük kestirilmesine yol açabilir.

Sireci, Thissen ve Wainer (1991) ikili puanlanan SAT verileri ile gerçekleştirdikleri çalışmalarında aynı madde takımında yer alan maddeler arası ilişkiyi dikkate almamanın hem KTK ya dayalı hem de MTK ya dayalı güvenilirlik kestiriminde %10-15 oranında yüksek tahmine yol açtığını belirlemiştir. Bu araştırmada da G kuramına dayalı hesaplanan G katsayısı yaklaşık olarak madde takımı etkisi dikkate alınmadığında, dikkate alındığı duruma göre %7 lik bir oranda yüksek elde edilmiştir. Madde takımları ile çalışıldığında daha doğru bir güvenilirlik kestirimi için madde takımı etkisinin ele alınması gerektiği sonucuna ulaşılmaktadır. Lee ve Park (2012), gerçekleştirdikleri simülasyon çalışmasında da bu araştırmanın sonuçlarına benzer şekilde G kuramı çerçevesinde madde takımı etkisinin dikkate alındığı durumda daha düşük güvenilirlik elde etmişlerdir. Aynı zamanda çalışmalarında G kuramı ve madde tepki kuramına dayalı analizleri gerçekleştirmişlerdir. G kuramının dikkate alındığı durumda daha düşük güvenilirlik değerlerini de raporlamışlardır. Yapılan K çalışmaları sonucunda çoktan seçmeli maddelerden oluşan başarı testlerinde, birey sayısı sabit tutulup madde ve madde takımı sayısı artırıldığında elde edilen güvenilirlik katsayılarında artış gözlenmiştir. Bu durum açık uçlu maddelerden oluşan testte olduğu gibi madde sayısının güvenirlığe etkisi ile açıklanabilmektedir. Yürütülen K çalışmaları görülmektedir ki birey sayısı sabit tutulduğunda madde sayısının artırılmasından ziyade madde takımı sayısının artırılması güvenirliliğin artmasında daha fazla katkı sağlayacaktır. Kaya Uyanık ve Gelbal (2018), gerçekleştirdikleri simülasyon çalışmalarında da bu çalışmanın sonuçlarına benzer şekilde madde takımlarının eşit olması durumunda madde sayısı arttığında daha yüksek güvenilirlik değeri elde etmişlerdir. Madde takımlarında yer alan madde sayısının eşit olmasında ise madde takımı sayısının artmasıyla daha yüksek güvenilirlik elde etmişlerdir. Kısacası madde sayısı toplamda arttığı durumlarda daha yüksek güvenilirlik elde edilmektedir.

Üçüncü Alt Probleme İlişkin Sonuç ve Tartışma

Benzer içeriğe sahip açık uçlu madde ve çoktan seçmeli maddelerden oluşan madde takımlarının yer aldığı başarı testlerinde elde edilen G ve Phi katsayıları madde takımı etkisinin ele alınmadığı ve dikkate alındığı durumlarda karşılaştırılmıştır. Madde takımı etkisinin ele alınmadığı durumda açık uçlu maddelerden oluşan testte daha yüksek G ve Phi katsayısı elde edilmiştir. Ancak madde takımı etkisi dikkate alındığında ise çoktan seçmeli maddelerden oluşan testte daha yüksek G ve Phi katsayısı elde edilmiştir. Madde takımı etkisi dikkate alındığında kestirilen güvenirlik katsayılarının madde takımı etkisi ele alınmadığı durumda kestirilen güvenirlik katsayılarından daha düşük olduğu görülmüştür. Açık uçlu maddelerden oluşan madde takımlarının yer aldığı başarı testinde madde takımı etkisinin dikkate alınması durumunda G katsayısı (0,14'lük fark) ve Phi katsayısında (0,12'lik fark) düşüşe neden olmuştur. Benzer farklılık, çoktan seçmeli maddelerden oluşan madde takımlarının yer aldığı başarı testindeki G katsayısı (0,05'lik fark) ve Phi katsayısı (0,05'lik fark) farklılığına kıyasla çok daha fazladır. Dolayısıyla G katsayısındaki değişimde, madde türlerinin etkili olduğu ifade edilebilir.

Çoktan seçmeli maddelerden oluşan madde takımlarının yer aldığı başarı testinden elde edilen G ve Phi katsayılarının hem madde takımı etkisinin ele alınmadığı hem de dikkate alındığı durumda birbirine çok yakın değerler aldığı (0,1'lik değişim) belirlenmiştir. Ancak açık uçlu maddelerden oluşan madde takımlarının yer aldığı başarı testinde ise genellenebilirlik kuramı kapsamında yapılan G çalışması sonucunda elde edilen G ve Phi katsayılarının çoktan seçmeli maddelerden oluşan teste kıyasla birbirinden biraz daha farklı (0,3 ve 0,2'lik değişim) olduğu sonucuna ulaşılmıştır.

Öneriler

Çalışma sonucunda elde edilen bulgulardan yola çıkarak ileride yapılacak olan başka çalışmalara ışık tutması ve araştırmacılara kolaylık sağlaması açısından sunulan öneriler aşağıda sıralanmıştır.

1. Madde takımlarından oluşan testlerin güvenilirlik kestirimleri yapılırken daha doğru sonuçlar elde etmek adına madde takımı etkisinin ele alındığı durumlarda güvenilirlik kestirimleri yapılması tercih edilmelidir.
2. Madde takımı içindeki madde sayısını artırmak yerine, madde takımı sayısını artırmak daha yüksek güvenilirlik kestirimi yapabilmek adına tercih edilebilir.
3. Bu araştırmada madde takımları içindeki madde sayıları eşit olduğundan çalışmalar dengelenmiş desenler üzerinde yürütülmüştür. Araştırmacılar bu çalışmayı her bir madde takımında bulunan madde sayılarının değişkenlik gösterdiği dengelenmemiş desenler üzerinde de çalışabilirler.
4. Çalışma 8. sınıf öğrencilerine matematik branşında uygulanmış olup farklı kademe ve branşlarda uygulaması yapılabilir.
5. Çalışmada bireylerin madde ve puanlayıcılarla çaprazlandığı desen ile maddelerin madde takımları içinde yuvalandığı ve bunların da birey ve puanlayıcılarla çaprazlandığı desenler kullanılmış olup araştırmacılar puanlayıcıların bireyler içinde ya da maddelerin puanlayıcılar içinde yuvalandığı farklı desenlerde çalışmalarını gerçekleştirebilirler.
6. Uygulama sırasında öğrencilerin açık uçlu maddelerden ziyade çoktan seçmeli maddelerden oluşan başarı testlerini cevaplandırmada daha istekli oldukları gözlenmiştir. Bu durumun göz önüne alınarak sınıf içi uygulamalarda açık uçlu maddelere daha çok yer verilmesi önerilmektedir. Bu sayede öğrencilerin bu madde formatını cevaplamaya daha istekli olmaları sağlanabilir.
7. Araştırmacılar madde takımlarının güvenilirliğe etkisini karma bir test hazırlayıp test dizaynı açısından inceleyebilirler. Bir testin ilk bölümünde açık uçlu maddelerden

oluřan madde takımları yer alırken dięer testin ilk bölümünde çoktan seçmeli maddelerden oluřan madde takımları yer alabilir. Böylece farklı madde formatında hazırlanan madde takımlarının sıra düzeninin güvenilirliğe etkisi incelenebilir.



KAYNAKLAR

- Akın Arıkan, Ç., & Gümüş Özyıldırım, F. (2020). Çoklu gösterimlerin kullanıldığı matematik problemlerine ait çözümlerin puanlayıcı güvenirliği açısından incelenmesi. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi*, 14(1), 610-628. doi: 10.17522/balikesirnef.687639
- Atılğan, H. (2019). *Genellenebilirlik kuramı ve uygulaması*. Ankara: Anı.
- Başokçu, T. O. (2019). Ölçme sonuçlarının nitelikleri: Ölçme hatası, güvenirlik, geçerlik ve kullanışlılık. N. Doğan (Ed.), *Eğitimde ölçme ve değerlendirme* içinde (s. 31-74). Ankara: Pegem Akademi.
- Başol, G., & Şevran, B. (2019). Bir genellenebilirlik kuramı analizi: Değerlendirmede kullanılan ölçme araçlarının güvenirliği “Haydi Elektriği İletelim”. *Journal of Human Sciences*, 16(1), 370-391. doi:10.14687/jhs.v16i1.5590
- Baykul, Y. (2000). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması*. Ankara: ÖSYM.
- Berberoğlu, G. (2006). *Sınıf içi ölçme değerlendirme teknikleri*. İstanbul: Morpa Kültür.
- Brennan, R. L. (2001). *Generalizability theory*. New-York: Springer-Verlag.
- Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory*, Ohio-USA: Cengage Learning.
- Çakan, M. (2004). Öğretmenlerin ölçme-değerlendirme uygulamaları ve yeterlik düzeyleri: İlk ve ortaöğretim. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 37(2), 99-114. <https://hdl.handle.net/20.500.12491/966> sayfasından erişilmiştir.

- Çakıcı Eser, D., & Gelbal, S. (2013). Genellenebilirlik kuramı ve lojistik regresyona dayalı hesaplanan puanlayıcılar arası tutarlığın karşılaştırılması. *Kastamonu Eğitim Dergisi*, 21(2), 421-438. <https://search.trdizin.gov.tr/tr/yayin/detay/147819/> sayfasından erişilmiştir.
- Çetin, B., & İlhan, M. (2017). Standart ve Solo taksonomisine dayalı rubrikler ile puanlanan açık uçlu matematik sorularında puanlayıcı katılımı ve cömertliğinin incelenmesi. *Eğitim ve Bilim*, 42(189), 217-247. doi: 10.15390/EB.2017.5082
- Demirbilek Zorba, İ. (2020). *Personel alımında kullanılan bir yazılı sınav sonucunun genellenebilirlik kuramındaki farklı desenlerle karşılaştırılması*. Yüksek Lisans Tezi, Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
- Doğan, N. (2009a). Yazılı yoklamalar. H. Atılgan (Ed.), *Eğitimde ölçme ve değerlendirme* içinde (s. 145-168). Ankara: Anı.
- Doğan, N. (2009b). Çoktan seçmeli testler. H. Atılgan (Ed.), *Eğitimde ölçme ve değerlendirme* içinde (s. 223-268). Ankara: Anı.
- Doğan, N. (2019). Geleneksel ölçme ve değerlendirme teknikleri: I Yanıtı seçmeyi gerektiren ölçme araçları. N. Doğan (Ed.), *Eğitimde ölçme ve değerlendirme* içinde (s. 113-138). Ankara: Pegem Akademi.
- Eckes, T. (2017). Rater effects: Advances in item response modeling of human ratings - Part I. *Psychological Test and Assessment Modeling*, 59(4), 443-452.
- Gessaroli, M. E., & Folske, J. C. (2002). Generalizing the reliability of tests comprised of testlets. *International Journal of Testing*, 2(3-4), 277-295. <https://doi.org/10.1080/15305058.2002.9669496>
- Goodrich, H. (1996). Understanding rubrics. *Educational Leadership*, 54(4), 14-17.
- Goodrich Andrade, H. (2000). Using rubrics to promote thinking and learning. *Educational Leadership*, 57(5), 13-18.

- Güler, N., & Gelbal, S. (2010). Açık uçlu matematik sorularının güvenilirliğinin klasik test kuramı ve genellenebilirlik kuramına göre incelenmesi. *Kuram ve Uygulamada Eğitim Bilimleri*, 10(2), 989-1019.
- Güler, N., Kaya Uyanık, G., & Taşdelen Teker, G. (2012). *Genellenebilirlik kuramı*. Ankara: Pegem Akademi.
- Güler, N., & Taşdelen Teker, G. (2015). Açık uçlu maddelerde farklı yaklaşımlarla elde edilen puanlayıcılar arası güvenilirliğin değerlendirilmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 6(1), 12-24. doi: 10.21031/epod.63041
- Haladyna, T. M. (2004). *Developing and validating multiple-choice test items*, Taylor & Francis Group, <https://ebookcentral.proquest.com/lib/gazi-ebooks/detail.action?docID=255610> sayfasından erişilmiştir.
- Hendrickson, A. B. (2001, April). *Reliability of scores from tests composed of testlets: A comparison of methods*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Seattle, WA.
- Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and Evaluation in Counseling and Development*, 34(3), 177-189. doi:10.1080/07481756.2002.12069034.
- Kamış, Ö., & Doğan, C. D. (2017). Genellenebilirlik kuramında gerçekleştirilen karar çalışmaları ne kadar kararlı? *Gazi Eğitim Fakültesi Dergisi*, 37(2), 591-610.
- Kan, A. (2009). Ölçme araçlarında bulunması gereken nitelikler. H. Atılgan (Ed.), *Eğitimde ölçme ve değerlendirme içinde* (s. 23-80). Ankara: Anı.
- Karasar, N. (1994). *Bilimsel Araştırma Yöntemi*. Ankara: 3A Araştırma Eğitim Danışmanlık.
- Kaya Uyanık, G., & Gelbal, S. (2018). Madde tepki modellemesinde genellenebilirlik ile iki yüzeyli desenlerin incelenmesi. *Journal of Measurement and Evaluation in Education and Psychology*, 9(1), 17-32. doi: 10.21031/epod.349718
- Kaya Uyanık, G., & Ertuna, L. (2022). Examination of testlet effect in open-ended items. *SAGE Open*, 1-12. <https://doi.org/10.1177/21582440221079849>

- Lee, G., & Frisbie, D. A. (1999). Estimating reliability under a generalizability theory model for test scores composed of testlets. *Applied Measurement in Education*, 12(3), 237–255. https://doi.org/10.1207/S15324818AME1203_2
- Lee, G., Brennan, R. L., & Frisbie, D. A. (2000). Incorporating the testlet concept in test score analyses. *Educational Measurement: Issues and Practice*, 19(4), 9-15. <https://doi.org/10.1111/j.1745-3992.2000.tb00041.x>
- Lee, G.; & Park, I.-Y. (2012). A comparison of the approaches of generalizability theory and item response theory in estimating the reliability of test scores for testlet-composed tests. *Asia Pacific Education Review*, 13(1), 47-54. doi: 10.1007/s12564-011-9170-0
- Miles, M. B. & Huberman, A. M. (1994). *Qualitative data analyses: An expanded sourcebook*. CA: Sage Publications.
- Moskal, B. M., & Leydens, J. A. (2000). Scoring rubric development: Validity and reliability. *Practical Assessment, Research and Evaluation*, 7(10), 1-6. doi: <https://doi.org/10.7275/q7rm-gg74>
- Özçelik, D. A. (2013). *Test hazırlama kılavuzu*. Ankara: Pegem Akademi.
- Saal, F. E., Downey, R. G., & Lahey, M. A. (1980). Rating the ratings: Assessing the psychometric quality of rating data. *Psychological Bulletin*, 88(2), 413-428. doi:10.1037/0033-2909.88.2.413
- Shavelson, R. J. & Webb, N. M. (1991). *Generalizability theory: a primer*. Sage Publications.
- Sireci, S. G.; Thissen, D., & Wainer, H. (1991). On the reliability of testlet-based tests. *Journal of Educational Measurement*, 28(3), 237-247. <https://www.jstor.org/stable/14314845> sayfasından erişilmiştir.
- Şata, M. (2019). *Performans değerlendirme sürecinde puanlayıcı eğitiminin puanlayıcı davranışları üzerindeki etkisinin incelenmesi*. Doktora Tezi, Gazi Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.

- Şata, M. (2022). Açık uçlu maddelerin analizi. İ. Karakaya (Ed.), *Açık uçlu soruların hazırlanması, uygulanması ve değerlendirilmesi* içinde (s. 110-113). Ankara: Pegem Akademi.
- Taşdelen Teker, G. (2014). *Madde takımlarının güvenilirlik ve değişen madde fonksiyonu üzerine etkisi*. Doktora Tezi, Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Taşdelen Teker, G., Şahin, M. G., & Baytemir, K. (2016). Using generalizability theory to investigate the reliability of peer assessment. *Journal of Human Sciences*, 13(3), 5574-5586. doi: 10.14687/jhs.v13i3.4155
- Taştan, Z., & Uzun, N. B. (2021). Genellenebilirlik kuramında çok yüzeyli desenlerin incelenmesi. *Türkiye Sosyal Araştırmalar Dergisi*, 25(3), 743-756. <https://dergipark.org.tr/tr/pub/tsadergisi/issue/66172/533575> sayfasından erişilmiştir.
- Tekin, H. (2009). *Eğitimde ölçme ve değerlendirme*. Ankara: Yargı.
- Thissen, D., Steinberg, L., & Mooney, J. A. (1989). Trace lines for testlets: A use of multiple-categorical-response models. *Journal of Educational Measurement*, 26(3), 247-260. doi: 10.1111/j.1745-3984.1989.tb00331.x
- Turgut, M. F. (1992). *Eğitimde ölçme ve değerlendirme metotları*. Ankara: Saydam.
- Turgut, M. F. & Baykul, Y. (2010). *Eğitimde ölçme ve değerlendirme*. Ankara: Pegem Akademi.
- Wainer, H., & Kiely, G. L. (1987). Item clusters and computerized adaptive testing: A case for testlets. *Journal of Educational Measurement*, 24(3), 185-201. <https://www.jstor.org/stable/1434630> sayfasından erişilmiştir.
- Wainer, H. (1995). Precision and differential item functioning on a testlet-based test: The 1991 law school admissions test as an example. *Applied Measurement in Education*, 8(2), 157-186. doi: 10.1207/s15324818ame0802_4

- Wainer, H., & Thissen, D. (1996). How is reliability related to the quality of test scores? What is the effect of local dependence on reliability? *Educational Measurement: Issues and Practice*, 15(1), 22-29. doi: 10.1111/j.1745-3992.1996tb00803.x
- Wainer, H., Bradlow, E. T., & Du, Z. (2000). Testlet response theory: An analog for the 3PL model useful in testlet-based adaptive testing. In W. J. van der Linden & G. A. W. Glas (Eds.), *Computerized adaptive testing: Theory and practice* (pp. 245–269). Springer. doi: 10.1007/0-306-47531-6_13
- Yaman, S. (2016). Çoktan seçmeli madde tipleri ve fen eğitiminde kullanılan örnekleri. *Gazi Eğitim Bilimleri Dergisi*, 2(2), 151-170. <https://dergipark.org.tr/tr/pub/gebd/issue/35205/390659> sayfasından erişilmiştir.

EKLER



**EK 1. Pilot Uygulama Sonucu Elde Edilen Çoktan Seçmeli Maddelerden Oluşan
Madde Takımlarının Yer Aldığı Başarı Testine Ait Tap Programı Analiz Çıktıları**

TITLE: ÇOKTAN SEÇMELİ PİLOY UYG.

COMMENT:

TITLE: ÇOKTAN SEÇMELİ PİLOY UYG.

COMMENT:

Quick Item Analysis

Item	Key	Number Correct	Item Diff	Disc. Index	# Correct in High Grp	# Correct in Low Grp	Point Biser	Adj PtBis
Item 01 (2)		58	0,50	0,52	29 (0,78)	12 (0,27)	0,47	0,34
Item 02 (2)		53	0,46	0,48	25 (0,68)	9 (0,20)	0,46	0,33
Item 03 (3)		49	0,43	0,45	25 (0,68)	10 (0,22)	0,38	0,25
Item 04 (2)		38	0,33	0,19	16 (0,43)	11 (0,24)	0,24	0,11
Item 05 (1)		67	0,58	0,73	37 (1,00)	12 (0,27)	0,62	0,52
Item 06 (3)		73	0,63	0,61	34 (0,92)	14 (0,31)	0,52	0,40
Item 07 (4)		42	0,37	0,66	27 (0,73)	3 (0,07)	0,57	0,46
Item 08 (4)		39	0,34	0,37	22 (0,59)	10 (0,22)	0,37	0,24
Item 09 (3)		48	0,42	0,39	26 (0,70)	14 (0,31)	0,43	0,30
Item 10 (3)#		21	0,18	0,16	10 (0,27)	5 (0,11)	0,19	0,08
Item 11 (1)#		17	0,15	0,12	7 (0,19)	3 (0,07)	0,14	0,04
Item 12 (1)		38	0,33	0,48	22 (0,59)	5 (0,11)	0,45	0,32
Item 13 (4)		55	0,48	0,54	29 (0,78)	11 (0,24)	0,48	0,35
Item 14 (3)		48	0,42	0,64	31 (0,84)	9 (0,20)	0,60	0,49
Item 15 (3)		41	0,36	0,52	25 (0,68)	7 (0,16)	0,55	0,44
Item 16 (1)		33	0,29	0,46	21 (0,57)	5 (0,11)	0,53	0,42

marks potential problems (p<0,2 or p>0,95, D<0, pbis<0, adjpbis<0)

These results have been sorted by item number

```

=====
Number of Items Excluded   = 0
Number of Items Analyzed  = 16
Mean Item Difficulty       = 0,391
Mean Discrimination Index  = 0,458
Mean Point Biserial       = 0,438
Mean Adj. Point Biserial   = 0,318
KR20 (Alpha)              = 0,730
KR21                      = 0,707
SEM (from KR20)           = 1,745
High Grp Min Score (n=37) = 8,000
Low Grp Max Score (n=45)  = 4,000
-----

```

```

# Potential Problem Items = 2
defined as: difficulty <= 0,20(2)
           or: difficulty >= 0,95(0)
           or: D index <= 0,00(0)
           or: AdjPtBiserial <= 0,00(0)

```

```

Total Possible Score= 16
Minimum Score       = 0,000 = 0,0%
Maximum Score       = 15,000 = 93,8%
Mean Score          = 6,261 = 39,1%
Standard Deviation  = 3,360

```

```

~~~~~
TAP: Test Analysis Program (version 19.1.4)
Copyright © 2003-2018 Gordon P. Brooks
Contact: brooksg@ohio.edu

```

```

-----
Number of Examinees = 115
Mean Score          = 6,261 = 39,1%
Standard Deviation  = 3,360

```

```

~~~~~
TAP: Test Analysis Program (version 19.1.4)
Copyright © 2003-2018 Gordon P. Brooks
Contact: brooksg@ohio.edu

```

EK 2. Çoktan Seçmeli Maddelerden Oluşan Madde Takımlarının Yer Aldığı Matematik Başarı Testinin Nihai Formu

Madde Takımı-1

Aşağıdaki bilgilere göre 1, 2 ve 3. soruları cevaplayınız.

Bir uçak firması kabin içi valizlerde 15 kilografa kadar izin vermekte, 15 kilogramın üzerine çıktığında ise ekstra ücret talep etmektedir. Bu uçak firması ile seyahat edecek olan Ali Bey'in yanına alacağı kabin içi valizin kütlesinin çözümlenmiş şekli

$$1.10^1 + 2.10^0 + 9.10^{-1} + 3.10^{-2} + 5.10^{-3} \text{ tür.}$$

Ali Bey'in valizine eklemek istediği eşyaların kütlesinin çözümlenmiş şekli aşağıda verilmiştir.

Almak İsteddiği Eşyalar	Eşyanın Kütlesi (kg)
Kitap	$4.10^{-1} + 3.10^{-3}$
Dergi	$6.10^{-1} + 8.10^{-2}$
Şampuan	$10^0 + 10^{-1} + 3.10^{-3}$
Ayakkabı	$8.10^{-1} + 5.10^{-2} + 1.10^{-3}$
Şemsiye	$4.10^{-1} + 2.10^{-3}$

Ali Bey ekstra ücret ödemek istemediğine göre;

- Aşağıdaki eşyalardan hangilerini valizine aynı anda ekleyemez?
A) Kitap – Dergi – Şemsiye
B) Dergi – Şampuan – Ayakkabı
C) Kitap – Şampuan – Şemsiye
D) Dergi – Ayakkabı – Şemsiye
- Kütleleri eşit olan kitaplardan en fazla kaç tane koyabilir?
A) 4 B) 5 C) 6 D) 7
- Yukarıda verilen eşyaların tümünü valizine koymak isterse ekstra ücret ödememek için valizinden kaç kilogram eşya çıkarmalıdır?
A) 0,374 B) 0,429 C) 1,374 D) 1,446

Madde Takımı-2

Aşağıdaki bilgilere göre 4, 5 ve 6. soruları cevaplayınız.

A B C Ç D E F G Ğ H I | J K L M N O Ö P R S Ş T U Ü V Y Z

Bora Öğretmen öğrencileri ile birlikte bir kodlama oyunu oynayacaktır. Verilen bir kelimenin kodunu bulmak için harflerin alfabedeki sırasına bakılacak ve harflerin bulunduğu sıradaki sayı tam kare ise karekökü alınacak, tam kare değilse karekökünün hangi iki tam sayı arasında olduğu belirlenip en yakın tam sayıya yuvarlanacaktır. Buldukları sayıları yan yana yazdıklarında kelimenin kodu oluşacaktır.

Örneğin Melek isminin kodu,

MELEK = $\sqrt{16} - \sqrt{6} - \sqrt{15} - \sqrt{6} - \sqrt{14} = 42424$ şeklindedir.

4. Buna göre YASİN isminin kodu nedir?

- A) 51534 B) 51434 C) 41534 D) 41434

5. Kodu 1515 olan isim aşağıdakilerden hangisi olabilir?

- A) Aral B) Asal C) Aras D) Ayan

6. Aylin bir kelimenin kodunu hesaplarken harfin alfabedeki sırasının karekökünün hangi iki tam sayı arasında olduğunu belirlemiştir. Ancak yuvarlama yaparken belirlediği bu tam sayılardan en yakın sayıya yuvarlamak yerine uzak olan sayıya yuvarlamıştır. Bu durumda bulduğu kod 3133 olduğuna göre Aylin'in bulmak istediği kodun doğrusu aşağıdakilerden hangisi olamaz?

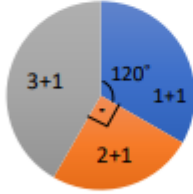
- A) 4124 B) 4142 C) 4144 D) 4244

Madde Takımı-3

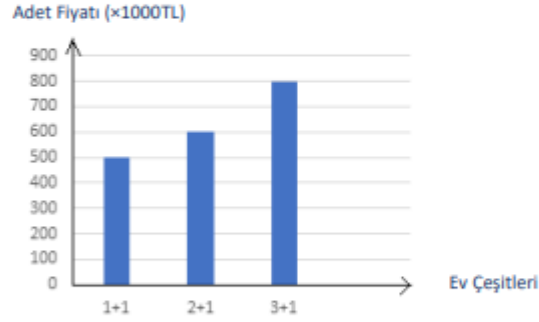
Aşağıdaki bilgilere göre 7, 8 ve 9. soruları cevaplayınız.

Bir sitede satışa sunulan evlerin oda sayılarına göre dağılımı daire grafiğinde, adet fiyatları ise sütun grafiğinde gösterilmiştir.

Evlerin Oda Sayılarına Göre Dağılımı



Evlerin Adet Fiyatları



7. Aşağıdakilerden hangisi bu sitede satışa sunulan 3+1 ev sayısı olamaz?

- A) 10 B) 150 C) 208 D) 300

8. Sitedeki evlerin tamamının satıldığı düşünülürse 1+1 ve 2+1 evlerin satışından elde edilen gelir farkı kaç TL olamaz?

- A) 200.000 B) 600.000 C) 750.000 D) 1.000.000

9. Sitede inşa edilen 2+1 ev sayısı 2 katına çıkarılırsa 2+1 ve 3+1 evlerin tamamının satışından elde edilen yeni gelir için aşağıdaki çıkarımlardan hangisi doğrudur?

- A) 2+1 evlerin satışından elde edilen gelir artmıştır.
B) 3+1 evlerin satışından elde edilen gelir artmıştır.
C) 2+1 evlerin satışından elde edilen gelir azalmıştır.
D) 3+1 evlerin satışından elde edilen gelir azalmıştır.

Madde Takımı-4

Aşağıdaki bilgilere göre 10, 11 ve 12. soruları cevaplayınız.

Aşağıda içerisinde toplar bulunan iki torba verilmiştir ve bu torbalardaki toplar renkleri dışında özdeşdir.



1. Torba



2. Torba

1. torbada sadece kırmızı ve beyaz renkli toplar bulunurken 2. torbada yalnız beyaz ve mavi renkli toplar vardır. Kırmızı top sayısı beyaz top sayısından fazla olan 1. torbada toplam 20 top, beyaz top sayısı mavi top sayısından fazla olan 2. torbada toplam 10 top bulunmaktadır.

10. 1. ve 2. torbadaki toplar karıştırılıp yeni bir torbaya atılıyor. Bu torbadan rastgele seçilecek bir topun beyaz olma olasılığı aşağıdakilerden hangisi olamaz?

A) $\frac{7}{30}$

B) $\frac{1}{3}$

C) $\frac{1}{2}$

D) $\frac{19}{30}$

11. 1. ve 2. torbadaki beyaz topların sayısının eşit, mavi topların sayısının ise maksimum olduğu bilinmektedir. Bu durumda iki torba karıştırılıp rastgele bir top seçildiğinde seçilen topun kırmızı olma olasılığı kaçtır?

A) $\frac{2}{15}$

B) $\frac{2}{5}$

C) $\frac{7}{15}$

D) $\frac{2}{3}$

12. 1.torbadaki kırmızı topların minimum 2.torbadaki mavi topların maksimum olduğu durumda iki torba karıştırılıp rastgele bir top seçiliyor. Seçilen topun beyaz olma olasılığı kaçtır?

A) $\frac{2}{15}$

B) $\frac{11}{30}$

C) $\frac{1}{2}$

D) $\frac{8}{15}$

2. Kütleleri eşit olan kitaplardan valizine en fazla kaç tane koyabilir? İşlem yaparak açıklayınız.

3. Yukarıda verilen eşyaların tümünü valizine koymak isterse kilo sınırını aşmamak için valizinden kaç kilogramlık eşya çıkarmalıdır? İşlem yaparak açıklayınız.

Madde Takımı-2

A B C Ç D E F G Ğ H I İ J K L M N O Ö P R S Ş T U Ü V Y Z

Bora Öğretmen öğrencileri ile birlikte bir kodlama oyunu oynayacaktır. Verilen bir kelimenin kodunu bulmak için harflerin alfabe'deki sırasına bakılacak ve harflerin bulunduğu sıradaki sayı tam kare ise karekökü alınacak, tam kare değilse hangi iki tam sayı arasında olduğu belirlenip en yakın tam sayıya yuvarlanacaktır. Buldukları sayıları yan yana yazdıklarında kelimenin kodu oluşacaktır.

Örneğin Gülru isminin kodu,

$$\text{GÜLRU} = \sqrt{8} - \sqrt{26} - \sqrt{15} - \sqrt{21} - \sqrt{25} = 35455 \text{ şeklindedir.}$$

Yukarıdaki bilgilere göre 4, 5 ve 6. soruları cevaplayınız.

4. Aşağıda verilen isimlerden bir tanesini seçip kodunu bulunuz.

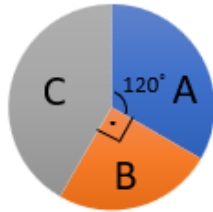
BEYZA AHMET KEREM MİRZA BERNA

5. Kodu 1515 olan bir isim bulunuz.

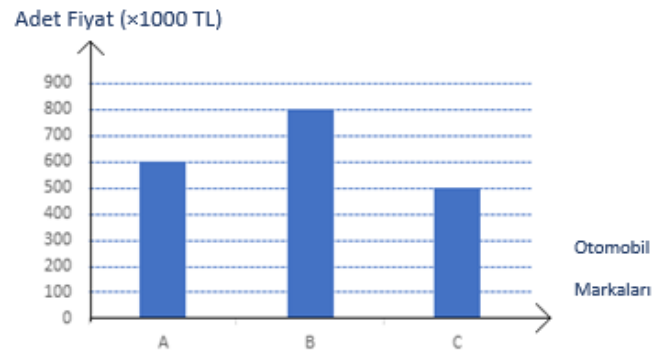
6. Aylin, bir kelimenin kodunu hesaplarken harfin alfabadeki sırasının karekökünün hangi iki tam sayı arasında olduğunu belirlemiştir. Ancak yuvarlama yaparken belirlediği bu tam sayılardan en yakın sayıya yuvarlamak yerine uzak olan sayıya yuvarlamıştır. Bu durumda bulduğu kod 3133 olduğuna göre Aylin'in bulmak istediği kodun doğrusu ne olabilir?

Bir fabrikada bir yılda üretilen otomobillerin sayılarına göre dağılımı daire grafiğinde, markalarına göre adet fiyatları sütun grafiğinde gösterilmiştir.

Bir fabrikada bir yılda üretilen otomobillerin sayılarına göre dağılımı daire grafiğinde, markalarına göre adet fiyatları sütun grafiğinde gösterilmiştir.



Otomobillerin Adet Fiyatları



Yukarıdaki bilgilere göre 7, 8 ve 9. soruları yanıtlayınız.

7. Bu fabrikada üretilen C otomobilinden bir yılda kaç adet üretilmiş olabilir? İşlem yaparak ifade ediniz.

8. Bir yılda üretilen otomobillerin tamamının satıldığı düşünülürse B ve C otomobillerinin satışından elde edilen gelir farkı kaç TL olabilir?

9. Bir yılda üretilen B marka otomobil sayısı 2 katına çıkarılırsa A ve B otomobillerinin tamamının satışından elde edilen yeni gelir farkı nasıl değişir, açıklayınız.

Madde Takımı-4

Aşağıda içerisinde toplar bulunan iki torba verilmiştir ve bu torbalardaki toplar renkleri dışında özdeşdir.



1. Torba



2. Torba

1. torbada sadece kırmızı ve beyaz renkli toplar bulunurken 2. torbada yalnız beyaz ve mavi renkli toplar vardır. Kırmızı top sayısı beyaz top sayısından fazla olan 1. torbada toplam 20 top, beyaz top sayısı mavi top sayısından fazla olan 2. Torbada ise toplam 10 top bulunmaktadır.

Yukarıdaki bilgilere göre 10, 11 ve 12. soruları yanıtlayınız.

10. 1. ve 2. torbadaki toplar karıştırılıp yeni bir torbaya atılıyor. Bu torbadan rastgele seçilecek bir topun beyaz olma olasılığı kaç olabilir? Açıklayınız.

11. 1. ve 2. torbadaki beyaz topların sayısının eşit, mavi topların sayısının ise maksimum olduğu durumda; iki torba karıştırılıp rastgele bir top seçildiğinde seçilen topun kırmızı olma olasılığını beyaz ve mavi top olma olasılığı ile karşılaştırınız.

12. 1.torbadaki kırmızı topların minimum 2.torbadaki mavi topların maksimum olduğu durumda iki torba karıştırılıp rastgele bir top seçiliyor. Seçilen topun kırmızı olma olasılığını bulup beyaz ve mavi top olma olasılığı ile karşılaştırınız.

EK 4. Açık Uçlu Maddelerin Yer Aldığı Başarı Testine Ait Dereceli Puanlama

Anahtarı

AÇIK UÇLU MADDE DERECELİ PUANLAMA ANAHTARI

Kodlar	Tam Doğru Yanıt	
10	1) Gül Hanım'ın Valizi: 12,325 kg Kitap: 0,502 kg; Tablet: 0,78 kg Makyaj çantası: 1,102 kg Ayakkabı: 0,842 kg Şemsiye:0,305 kg 10 farklı seçim yapılabilir: Kitap – tablet – makyaj çantası Kitap – tablet – ayakkabı – şemsiye (3'lü kombinasyonları olabilir; KTA, KTŞ, KAŞ, TAŞ) Kitap – makyaj çantası – ayakkabı Kitap – makyaj çantası – şemsiye Makyaj çantası – ayakkabı – şemsiye Tablet – şemsiye – makyaj çantası 5 ve yukarısını bulanlar doğru kabul edilecektir.	3 Puan
	Kısmen Doğru Yanıtlar	
20	İşlem yapılmış olup 3 ve 5 arası sonuç bulunması	2 Puan
21	3'ten az seçim yapıldığı belirtilip, işlem yapılmaması	1 Puan
	Yanlış Yanıtlar	
30	Ağırlığı 15 kg'ı aşacak şekilde eşya seçimi yapılması	0 Puan
40	Boş	0 Puan
	Diğer Yanıtlar	
50	Soruyla ilgili olmayan bir eşyanın yazılması	0 Puan

Kodlar	Tam Doğru Yanıt	
10	2) Gül Hanım'ın Valizi: 12,325 kg Kitap: 0,502 kg $15 - 12,325 = 2,675$ kg $2,675 : 0,502 = 5,3$ En fazla 5 adet kitap koyabilir.	3 Puan

	Kısmen Doğru Yanıtlar	
20	İşlemi doğrudur ama farklı sayıda bir kitap yazmıştır. Örneğin 5,3'ü bulup 6 kitap koyabilir gibi yazması. Çıkarma veya bölme işleminde hata yapıp yine de 5 kitap koyabilir diyenler.	2 Puan
21	Sadece 5 kitap koyabilir deyip işlem yapmayanlar veya kitabın veya çantanın ağırlığını yanlış bulup ona göre doğru cevap bulanlar	1 Puan
	Yanlış Yanıtlar	
30	5'ten farklı bir rakam yazıp açıklaması olmayanlar	0 Puan
40	Boş	0 Puan
	Diğer Yanıtlar	
50	Soruyla ilgili olmayan bir şeyin yazılması	0 Puan

Kodlar	Tam Doğru Yanıt	
10	3) Gül Hanım'ın Valizi: 12,325 kg Kitap: 0,502 kg; Tablet: 0,78 kg Makyaj çantası: 1,102 kg Ayakkabı: 0,842 kg Şemsiye:0,305 kg $12,325 + 3,531 = 15,856$ kg Valizinden 0,856 kg'lık eşya çıkarmalıdır.	3 Puan
	Kısmen Doğru Yanıtlar	
20	İşlemin doğru olması fakat toplama veya çıkarmada işlem hatası yapılarak farklı bir değer bulunması.	2 Puan
21	Sadece 0,856 kg yazıp işlemin olmaması durumu ya da sadece 3,531'i bulması	1 Puan
	Yanlış Yanıtlar	
30	Herhangi bir açıklama olmadan farklı bir değer yazılması.	0 Puan
40	Boş	0 Puan
	Diğer Yanıtlar	
50	Soruyla ilgili olmayan bir şeyin yazılması	0 Puan
Kodlar	Tam Doğru Yanıt	
10	4) BEYZA=$\sqrt{2} - \sqrt{6} - \sqrt{28} - \sqrt{29} - \sqrt{1} = 12551$ AHMET=$\sqrt{1} - \sqrt{10} - \sqrt{16} - \sqrt{6} - \sqrt{24} = 13425$ KEREM=$\sqrt{14} - \sqrt{6} - \sqrt{21} - \sqrt{6} - \sqrt{16} = 42524$	3 Puan

	$MİRZA = \sqrt{16} - \sqrt{12} - \sqrt{21} - \sqrt{29} - \sqrt{1} = 43551$ $BERNA = \sqrt{2} - \sqrt{6} - \sqrt{21} - \sqrt{17} - \sqrt{1} = 12541$ Herhangi birini doğru yapanlar tam puan alacaktır.	
	Kısmen Doğru Yanıtlar	
20	Kodu bulurken harfin sırasını yanlış yazıp ona göre kod bulanlar veya Sadece bir harfin kodunda hata yapanlar	2 Puan
21	İki harfin kodunda hata yapanlar	1 Puan
	Yanlış Yanıtlar	
30	Kodu yanlış bulanlar İkiden fazla harfin kodunda hata yapanlar	0 Puan
40	Boş	0 Puan
	Diğer Yanıtlar	
50	Yukarıda verilen çözümle ilgili olmayan yanıtlar	0 Puan

Kodlar	Tam Doğru Yanıt	
10	5) $ARAS = \sqrt{1} - \sqrt{21} - \sqrt{1} - \sqrt{22} = 1515$ $AYAZ = \sqrt{1} - \sqrt{28} - \sqrt{1} - \sqrt{29} = 1515$ Kodu 1515 olan isimler kabul edilecektir.	3 Puan
	Kısmen Doğru Yanıtlar	
20	Kodu doğru bulup, ismi yazarken hata yapanlar (kök içindeki sıra numaraları doğru fakat ismi yanlış yazmış olabilir)	2 Puan
21	Kodu bulurken harfin sırasını yanlış yazıp ona göre isim bulanlar Anlamlı olmayan ama koda göre doğru kelime yazanlar	1 Puan
	Yanlış Yanıtlar	

30	Yanlış isim bulanlar	0 Puan
40	Boş	0 Puan
	Diğer Yanıtlar	
50	Yukarıda verilen çözümle ilgili olmayan yanıtlar	0 Puan

Kodlar	Tam Doğru Yanıt	
10	6) 3133 yerine 4142, 4144, 4124 gibi kodlar olabilir. Bir tanesini yazan doğru kabul edilecektir.	3 Puan
	Kısmen Doğru Yanıtlar	
20	2122, 2124, 2142, 4122, 2144 kodları ile ilgili anlamlı bir kelime olmadığı için bunları yazanların cevabı kısmen doğru kabul edilecektir.	2 Puan
21	Kodu bulurken sadece bir sayıda hata yapanlar	1 Puan
	Yanlış Yanıtlar	
30	Kodu yanlış bulanlar	0 Puan
40	Boş	0 Puan
	Diğer Yanıtlar	
50	Soru ile ilgili olmayan yanıtlar	0 Puan

Kodlar	Tam Doğru Yanıt	
10	7) A B C 120° 90° 150° 4x 3x 5x C marka araç sayısı 5'in katı olmalıdır. Bu durumda en az 5 araç üretilmiş olup, 5'in katı olan bütün sayılar doğru cevap kabul edilir. Derece sayısı işlem kolaylığı açısından direkt araç sayısı olarak kabul edilebilir. Bu açıklamada yeterli kabul edilecektir. (150 yazanlar tam puan alacaktır.	3 Puan
	Kısmen Doğru Yanıtlar	
20	C'nin derecesini yanlış bulup ona göre işlemi doğru yapanlar	2 Puan
21	150'nin katı olan sayılar / 5 in katı olan sayılar (açıklama yok)	1 Puan
	Yanlış Yanıtlar	
30	5'in ya da 150'nin katı olmayan sayılar	0 Puan

40	Boş	0 Puan
	Diğer Yanıtlar	
50	Soru ile ilgili olmayanlar cevaplar	0 Puan

Kodlar	Tam Doğru Yanıt	
10	8) A B C 120° 90° 150° 4x 3x 5x B marka araç satışından elde edilen gelir $3x \cdot 800\,000 = 2\,400\,000x$ C marka araç satışından elde edilen gelir $5x \cdot 500\,000 = 2\,500\,000x$ Gelir farkı = $100\,000x$ Bu durumda gelir farkı en az 100 000 TL ya da katları olacaktır.	3 Puan
	Kısmen Doğru Yanıtlar	
20	Sadece 100 000 TL veya katını yazıp yukarıdakine benzer bir açıklama yapmayanlar	2 Puan
21	Açıklaması doğru olup işlem hatası yapanlar	1 Puan
	Yanlış Yanıtlar	
30	100 000 ve katı olmayan sayılar	0 Puan
40	Boş	0 Puan
	Diğer Yanıtlar	
50	Soru ile ilgili olmayanlar cevaplar	0 Puan

Kodlar	Tam Doğru Yanıt	
10	9) A B C 96° 144° 120° 4x 6x 5x İlk durumda; B marka araç satışından elde edilen gelir $3x \cdot 800\,000 = 2\,400\,000x$ A marka araç satışından elde edilen gelir $4x \cdot 600\,000 = 2\,400\,000x$ Gelir farkı = 0 TL	

	İkinci durumda; B marka araç satışından elde edilen gelir $6x \cdot 800\,000 = 4\,800\,000x$ A marka araç satışından elde edilen gelir $4x \cdot 600\,000 = 2\,400\,000x$ Gelir farkı = $2\,400\,000x$ TL Buna göre, ilk durumda gelirler eşit iken; son durumda $2\,400\,000x$ TL B marka aracın lehine olmuştur.	
	Kısmen Doğru Yanıtlar	
20	Sadece $2\,400\,000$ TL veya katını yazıp yukarıdakine benzer bir açıklama yapmayanlar	2 Puan
21	Açıklaması doğru olup işlem hatası yapanlar	1 Puan
	Yanlış Yanıtlar	
30	Yanlış gelir farkı bulanlar	0 Puan
40	Boş	0 Puan
	Diğer Yanıtlar	
50	Soru ile ilgili olmayanlar cevaplar	0 Puan

Kodlar	Tam Doğru Yanıt	
10	10. 1.torbadaki beyaz toplar 1'den 9'a kadar değer alabilirken, 2.torbadaki beyaz toplar 6'dan 9'a kadar değer alabilir. Toplam top sayısı 30 olacaktır. Beyaz topların sayısı ise; 1-6 2-6.....9-6 1-7 2-7.....9-7 1-8 2-8.....9-8 1-9 2-9.....9-9 olabilir. Toplamda en düşük beyaz top sayısı 7, en fazla beyaz top sayısı 18 olacağından olasılık değeri $7/30$ ile $18/30$ arasında değer alabilir (Bu değerler de dahildir). Beyaz top sayısına yukarıdaki açıklamaya uygun olarak istediği şekilde karar verip bu sonuçlardan herhangi birini bulanlar tam puan alacaktır.	3 Puan
	Kısmen Doğru Yanıtlar	
20	Açıklaması doğru olup olasılık değerini yanlış yazanlar	2 Puan
21	Olasılık değerini doğru bulup açıklama yapmış ama açıklaması yanlış olanlar ya da hiç açıklaması olmayanlar	1 Puan
	Yanlış Yanıtlar	
30	Mavi ya da kırmızı topların olasılık değerini bulanlar (yanlış değerler)	0 Puan

40	Boş	0 Puan
	Diğer Yanıtlar	
50	Soru ile ilgili olmayan yanıtlar	0 Puan

Kodlar	Tam Doğru Yanıt	
10	11. 2.torbada maksimum 4 adet mavi top olabilir. Bu durumda beyaz top sayısı her iki torbada da 6 olacaktır. Kırmızı top sayısı ise 14 olur. Kırmızı olma olasılığı: 14/30 Mavi olma olasılığı: 4/30 Beyaz olma olasılığı: 12/30 olup kırmızı olma olasılığı diğer toplara göre daha fazla olur. (Kırmızı olma olasılığını mavi ve beyaz olma olasılığının toplamı ile karşılaştıranlar da tam puan alacaktır.)	3 Puan
	Kısmen Doğru Yanıtlar	
20	Bütün topların olasılıklarını doğru hesaplamış fakat karşılaştırma yapmamıştır ya da sonucu etkilemeyecek bir işlem hatası yapmıştır.	2 Puan
21	En fazla iki renk topun olma olasılıklarını hesaplamıştır.	1 Puan
	Yanlış Yanıtlar	
30	Top sayılarını ve olasılıklarını yanlış hesaplamıştır.	0 Puan
40	Boş	0 Puan
	Diğer Yanıtlar	
50	Soru ile ilgili olmayan yanıtlar	0 Puan

Kodlar	Tam Doğru Yanıt	
10	12. 1.torbada 11 kırmızı 9 beyaz top olur. 2.torbada ise 6 beyaz 4 mavi top vardır. Bu durumda; Mavi top olma olasılığı 4/30, Beyaz top olma olasılığı 15/30 Kırmızı top olma olasılığı 11/30 olup; kırmızı top olma olasılığı beyaz top olma olasılığından az mavi top olma olasılığından fazladır. (Kırmızı olma olasılığını mavi ve beyaz olma olasılığının toplamı ile karşılaştıranlar da tam puan alacaktır.)	3 Puan

	Kısmen Doğru Yanıtlar	
20	Olasılık değerlerini doğru bulup ters açıklama yapanlar ya da karşılaştırma yapmayanlar	2 Puan
21	Sadece “Kırmızı top olma olasılığı beyaz top olma olasılığından daha az mavi top olma olasılığından daha fazladır.” diyenler	1 Puan
	Yanlış Yanıtlar	
30	Olasılık değerlerini yanlış bulanlar veya yanlış açıklama yapanlar	0 Puan
40	Boş	0 Puan
	Diğer Yanıtlar	
50	Soru ile ilgili olmayan yanıtlar	0 Puan



**EK 5. Nihai Uygulama Sonucu Elde Edilen Çoktan Seçmeli Maddelerden Oluşan
Madde Takımlarının Yer Aldığı Başarı Testine Ait TAP Programı Analiz Çıktıları**

TITLE: ÇSS_128

COMMENT:

TITLE: ÇSS_128

COMMENT:

Quick Item Analysis

Item	Key	Number Correct	Item Diff	Disc. Index	# Correct in High Grp	# Correct in Low Grp	Point Biser	Adj PtBis
Item 01	(2)	59	0,46	0,59	34 (0,83)	10 (0,24)	0,57	0,44
Item 02	(2)	55	0,43	0,59	31 (0,76)	7 (0,17)	0,56	0,43
Item 03	(3)	56	0,44	0,44	26 (0,63)	8 (0,20)	0,39	0,24
Item 04	(1)	91	0,71	0,59	41 (1,00)	17 (0,41)	0,51	0,39
Item 05	(3)	76	0,59	0,68	39 (0,95)	11 (0,27)	0,55	0,42
Item 06	(4)	41	0,32	0,49	24 (0,59)	4 (0,10)	0,53	0,40
Item 07	(3)	37	0,29	0,51	24 (0,59)	3 (0,07)	0,50	0,38
Item 08	(3)	46	0,36	0,54	28 (0,68)	6 (0,15)	0,49	0,36
Item 09	(1)	47	0,37	0,63	30 (0,73)	4 (0,10)	0,60	0,48
Item 10	(4)	48	0,38	0,71	33 (0,80)	4 (0,10)	0,63	0,52
Item 11	(3)	65	0,51	0,71	36 (0,88)	7 (0,17)	0,57	0,44
Item 12	(3)	58	0,45	0,46	27 (0,66)	8 (0,20)	0,45	0,30
=====								
These results have been sorted by item number								
=====								

Number of Items Excluded = 0
Number of Items Analyzed = 12
Mean Item Difficulty = 0,442
Mean Discrimination Index = 0,577
Mean Point Biserial = 0,528
Mean Adj. Point Biserial = 0,401
KR20 (Alpha) = 0,765
KR21 = 0,747
SEM (from KR20) = 1,485
High Grp Min Score (n=41) = 7,000
Low Grp Max Score (n=41) = 3,000

Total Possible Score= 12
Minimum Score = 0,000 = 0,0%
Maximum Score = 12,000 = 100,0%
Mean Score = 5,305 = 44,2%
Standard Deviation = 3,063

~~~~~  
TAP: Test Analysis Program (version 19.1.4)  
Copyright © 2003-2018 Gordon P. Brooks  
Contact: brooksg@ohio.edu

-----  
Number of Examinees = 128  
Mean Score = 5,305 = 44,2%  
Standard Deviation = 3,063

~~~~~  
TAP: Test Analysis Program (version 19.1.4)
Copyright © 2003-2018 Gordon P. Brooks
Contact: brooksg@ohio.edu

EK 6. EduG Programı Analiz Çıktıları

Çoktan Seçmeli Maddelerden Oluşan Başarı Testinde Madde Takımı Etkisinin Ele Alınmadığı $b \times m$ Desenine Ait Analiz Çıktıları



Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
birey	B	128	INF	
madde	M	12	INF	

Analysis of variance

Source	SS	df	MS	Components					SE
				Random	Mixed	Corrected	%		
B	100.093	127	0.788	0.050	0.050	0.050	20.2		0.008
M	19.898	11	1.809	0.013	0.013	0.013	5.1		0.006
BM	258.852	1397	0.185	0.185	0.185	0.185	74.7		0.007
Total	378.843	1535					100%		

G Study Table (Measurement design B/M)

Source of variance	Differentiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
B	0.050					
		M			0.001	6.4
		BM	0.015	100.0	0.015	93.6
Sum of variances	0.050		0.015	100%	0.016	100%
Standard deviation	0.224		Relative SE: 0.124		Absolute SE: 0.128	
Coef_G relative	0.76					
Coef_G absolute	0.75					

Grand mean for levels used: 0.442
Variance error of the mean for levels used: 0.002
Standard error of the grand mean: 0.040

Optimization



	G-study		Option 1		Option 2		Option 3		Option 4		Option 5	
	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.
B	128	INF	128	INF	128	INF	128	INF	128	INF	128	INF
M	12	INF	6	INF	9	INF	15	INF	18	INF	21	INF
Observ.		1536		768		1152		1920		2304		2688
Coef_G rel.		0.765		0.619		0.709		0.803		0.830		0.851
rounded		0.76		0.62		0.71		0.80		0.83		0.85
Coef_G abs.		0.753		0.604		0.695		0.792		0.820		0.842

rounded	0.75	0.60	0.70	0.79	0.82	0.84
Rel. Err. Var.	0.015	0.031	0.021	0.012	0.010	0.009
Rel. Std. Err. of M.	0.124	0.176	0.143	0.111	0.101	0.094
Abs. Err. Var.	0.016	0.033	0.022	0.013	0.011	0.009
Abs. Std. Err. of M.	0.128	0.182	0.148	0.115	0.105	0.097

Çoktan Seçmeli Maddelerden Oluşan Başarı Testinde Madde Takımı Etkisinin Dikkate Alındığı $b \times (m:t)$ Desenine Ait Analiz Çıktıları

CSS TESTLET 128

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
birey	B	128	INF	
testlet	H	4	INF	
madde	M:H	3	INF	

Analysis of variance

Source	SS	df	MS	Components				
				Random	Mixed	Corrected	%	SE
B	100.093	127	0.788	0.047	0.047	0.047	18.8	0.008
H	7.929	3	2.643	0.003	0.003	0.003	1.1	0.005
M:H	11.969	8	1.496	0.010	0.010	0.010	4.2	0.005
BH	86.154	381	0.226	0.019	0.019	0.019	7.5	0.006
BM:H	172.698	1016	0.170	0.170	0.170	0.170	68.3	0.008
Total	378.843	1535					100%	

G Study Table
(Measurement design B/HM)

Source of variance	Differentiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
B	0.047					
		H			0.001	3.5
		M:H			0.001	4.2
		BH	0.005	24.8	0.005	22.9
		BM:H	0.014	75.2	0.014	69.4
Sum of variances	0.047		0.019	100%	0.020	100%
Standard deviation	0.216		Relative SE:	0.137	Absolute SE:	0.143
Coef_G relative	0.71					
Coef_G absolute	0.70					

Grand mean for levels used: 0.442

Variance error of the mean for levels used: 0.002

Standard error of the grand mean: 0.046

Optimization

	G-study		Option 1		Option 2		Option 3		Option 4		Option 5	
	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.
B	128	INF	128	INF	128	INF	128	INF	128	INF	128	INF
H	4	INF	4	INF	4	INF	4	INF	4	INF	4	INF
M:H	3	INF	2	INF	4	INF	5	INF	6	INF	7	INF
Observ.		1536		1024		2048		2560		3072		3584
Coef_G rel.		0.713		0.644		0.754		0.780		0.799		0.813
rounded		0.71		0.64		0.75		0.78		0.80		0.81
Coef_G abs.		0.696		0.626		0.738		0.765		0.784		0.798
rounded		0.70		0.63		0.74		0.76		0.78		0.80
Rel. Err. Var.		0.019		0.026		0.015		0.013		0.012		0.011
Rel. Std. Err. of M.		0.137		0.161		0.124		0.115		0.108		0.104
Abs. Err. Var.		0.020		0.028		0.017		0.014		0.013		0.012
Abs. Std. Err. of M.		0.143		0.167		0.129		0.120		0.114		0.109

Optimization

	G-study		Option 1		Option 2		Option 3		Option 4		Option 5	
	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.
B	128	INF	128	INF	128	INF	128	INF	128	INF	128	INF
H	4	INF	2	INF	3	INF	5	INF	6	INF	7	INF

M:H	3	INF	3	INF	3	INF	3	INF	3	INF	3	INF
Observ.	1536	768	1152	1920	2304	2688						
Coef. G rel.	0.713	0.554	0.651	0.756	0.788	0.813						
rounded	0.71	0.55	0.65	0.76	0.79	0.81						
Coef. G abs.	0.696	0.534	0.632	0.741	0.775	0.801						
rounded	0.70	0.53	0.63	0.74	0.77	0.80						
Rel. Err. Var.	0.019	0.038	0.025	0.015	0.013	0.011						
Rel. Std. Err. of M.	0.137	0.194	0.159	0.123	0.112	0.104						
Abs. Err. Var.	0.020	0.041	0.027	0.016	0.014	0.012						
Abs. Std. Err. of M.	0.143	0.202	0.165	0.128	0.117	0.108						

Açık Uçlu Maddelerden Oluşan Başarı Testinde Madde Takımı Etkisinin Ele Alınmadığı $b \times m \times p$ Desenine Ait Analiz Çıktıları

ana uyg aus 128

Observation and Estimation Designs



Facet	Label	Levels	Univ.	Reduction (levels to exclude)
birey	B	128	INF	
puanlayıcı	P	3	INF	
madde	M	12	INF	

Analysis of variance

Source	SS	df	MS	Components				
				Random	Mixed	Corrected	%	SE
B	1352.284	127	10.648	0.238	0.238	0.238	19.3	0.037
P	7.461	2	3.731	0.000	0.000	0.000	0.0	0.002
M	551.671	11	50.152	0.115	0.115	0.115	9.3	0.051
BP	91.816	254	0.361	0.007	0.007	0.007	0.6	0.003
BM	2773.607	1397	1.985	0.571	0.571	0.571	46.2	0.025
PM	92.247	22	4.193	0.031	0.031	0.031	2.5	0.009
BPM	763.809	2794	0.273	0.273	0.273	0.273	22.1	0.007
Total	5632.895	4607					100%	

G Study Table (Measurement design B/PM)

Source of variance	Differentiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
B	0.238					
		P			(0.000)	0.0
		M			0.010	14.1
		BP	0.002	4.2	0.002	3.6
		BM	0.048	82.6	0.048	69.9
		PM			0.001	1.2
		BPM	0.008	13.2	0.008	11.2
Sum of variances	0.238		0.058	100%	0.068	100%
Standard deviation	0.488		Relative SE: 0.240		Absolute SE: 0.261	
Coef_G relative	0.81					
Coef_G absolute	0.78					

Grand mean for levels used: 0.644

Variance error of the mean for levels used: 0.013

Standard error of the grand mean: 0.113

Optimization

	G-study		Option 1		Option 2		Option 3		Option 4		Option 5	
	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.
B	128	INF	128	INF	128	INF	128	INF	128	INF	128	INF
P	3	INF	3	INF	3	INF	3	INF	3	INF	3	INF
M	12	INF	6	INF	9	INF	15	INF	18	INF	21	INF
Observ.		4608		2304		3456		5760		6912		8064
Coef_G rel.		0.805		0.679		0.758		0.836		0.859		0.875
rounded		0.81		0.68		0.76		0.84		0.86		0.88
Coef_G abs.		0.778		0.641		0.726		0.813		0.838		0.856
rounded		0.78		0.64		0.73		0.81		0.84		0.86
Rel. Err. Var.		0.058		0.113		0.076		0.047		0.039		0.034
Rel. Std. Err. of M.		0.240		0.336		0.276		0.216		0.198		0.184
Abs. Err. Var.		0.068		0.134		0.090		0.055		0.046		0.040
Abs. Std. Err. of M.		0.261		0.366		0.300		0.234		0.215		0.200

Optimization

G-study	Option 1	Option 2	Option 3	Option 4	Option 5
---------	----------	----------	----------	----------	----------

	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.
B	128	INF	128	INF	128	INF	128	INF	128	INF	128	INF
P	3	INF	2	INF	4	INF	5	INF	6	INF	7	INF
M	12	INF	12	INF	12	INF	12	INF	12	INF	12	INF
Observ.		4608		3072		6144		7680		9216		10752
Coef_G rel.		0.805		0.792		0.812		0.816		0.819		0.821
rounded		0.81		0.79		0.81		0.82		0.82		0.82
Coef_G abs.		0.778		0.764		0.785		0.789		0.792		0.794
rounded		0.78		0.76		0.78		0.79		0.79		0.79
Rel. Err. Var.		0.058		0.063		0.055		0.054		0.053		0.052
Rel. Std. Err. of M.		0.240		0.250		0.235		0.231		0.229		0.228
Abs. Err. Var.		0.068		0.073		0.065		0.064		0.063		0.062
Abs. Std. Err. of M.		0.261		0.271		0.256		0.252		0.250		0.249

Açık Uçlu Maddelerden Oluşan Başarı Testinde Madde Takımı Etkisinin Dikkate Alındığı $b \times (m:t) \times p$ Desenine Ait Analiz Çıktıları

ana uyg aus mte 128

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
birey	B	128	INF	
puanlayıcı	P	3	INF	
testlet	H	4	INF	
madde	M:H	3	INF	

Analysis of variance

Source	SS	df	MS	Components				
				Random	Mixed	Corrected	%	SE
B	1352.284	127	10.648	0.199	0.199	0.199	16.0	0.037
P	7.461	2	3.731	-0.001	-0.001	-0.001	0.0	0.002
H	145.362	3	48.454	-0.005	-0.005	-0.005	0.0	0.033
M:H	406.309	8	50.789	0.119	0.119	0.119	9.6	0.059
BP	91.816	254	0.361	0.003	0.003	0.003	0.3	0.003
BH	1318.138	381	3.460	0.218	0.218	0.218	17.6	0.029
BM:H	1455.469	1016	1.433	0.392	0.392	0.392	31.6	0.021
PH	30.228	6	5.038	0.003	0.003	0.003	0.2	0.007
PM:H	62.019	16	3.876	0.028	0.028	0.028	2.3	0.010
BPH	243.605	762	0.320	0.021	0.021	0.021	1.7	0.006
BPM:H	520.203	2032	0.256	0.256	0.256	0.256	20.6	0.008
Total	5632.895	4607					100%	

G Study Table
(Measurement design B/PHM)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
B	0.199					
		P			(0.000)	0.0
		H			(0.000)	0.0
		M:H			0.010	9.2
		BP	0.001	1.2	0.001	1.1
		BH	0.055	56.1	0.055	50.4
		BM:H	0.033	33.6	0.033	30.2
		PH			0.000	0.2
		PM:H			0.001	0.7
		BPH	0.002	1.8	0.002	1.6
		BPM:H	0.007	7.3	0.007	6.6
Sum of variances	0.199		0.097	100%	0.108	100%
Standard deviation	0.446		Relative SE:	0.312	Absolute SE:	0.329
Coef_G relative	0.67					

Coef_G absolute 0.65

Grand mean for levels used: 0.644

Variance error of the mean for levels used: 0.013

Standard error of the grand mean: 0.115

Optimization

	G-study		Option 1		Option 2		Option 3		Option 4		Option 5	
	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.
B	128	INF	128	INF	128	INF	128	INF	128	INF	128	INF
P	3	INF	3	INF	3	INF	3	INF	3	INF	3	INF
H	4	INF	4	INF	4	INF	4	INF	4	INF	4	INF
M:H	3	INF	1	INF	2	INF	4	INF	5	INF	6	INF
Observ.		4608		1536		3072		6144		7680		9216
Coef_G rel.		0.671		0.529		0.629		0.695		0.709		0.720
rounded		0.67		0.53		0.63		0.69		0.71		0.72
Coef_G abs.		0.647		0.487		0.598		0.675		0.693		0.705
rounded		0.65		0.49		0.60		0.67		0.69		0.71
Rel. Err. Var.		0.097		0.177		0.117		0.087		0.081		0.077
Rel. Std. Err. of M.		0.312		0.421		0.342		0.295		0.285		0.278
Abs. Err. Var.		0.108		0.209		0.133		0.096		0.088		0.083
Abs. Std. Err. of M.		0.329		0.457		0.365		0.309		0.297		0.288

Optimization

	G-study		Option 1		Option 2		Option 3		Option 4		Option 5	
	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.
B	128	INF	128	INF	128	INF	128	INF	128	INF	128	INF
P	3	INF	3	INF	3	INF	3	INF	3	INF	3	INF
H	4	INF	3	INF	5	INF	6	INF	7	INF	8	INF
M:H	3	INF	3	INF	3	INF	3	INF	3	INF	3	INF
Observ.	4608		3456		5760		6912		8064		9216	
Coef_G rel.	0.671		0.606		0.718		0.753		0.780		0.801	
rounded	0.67		0.61		0.72		0.75		0.78		0.80	
Coef_G abs.	0.647		0.580		0.696		0.732		0.761		0.784	
rounded	0.65		0.58		0.70		0.73		0.76		0.78	
Rel. Err. Var.	0.097		0.129		0.078		0.065		0.056		0.049	
Rel. Std. Err. of M.	0.312		0.360		0.279		0.255		0.237		0.222	
Abs. Err. Var.	0.108		0.144		0.087		0.073		0.062		0.055	
Abs. Std. Err. of M.	0.329		0.379		0.295		0.269		0.250		0.234	

Optimization

	G-study		Option 1		Option 2		Option 3		Option 4		Option 5	
	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.
B	128	INF	128	INF	128	INF	128	INF	128	INF	128	INF
P	3	INF	2	INF	4	INF	5	INF	6	INF	7	INF
H	4	INF	4	INF	4	INF	4	INF	4	INF	4	INF
M:H	3	INF	3	INF	3	INF	3	INF	3	INF	3	INF
Observ.	4608		3072		6144		7680		9216		10752	
Coef_G rel.	0.671		0.660		0.677		0.680		0.683		0.684	
rounded	0.67		0.66		0.68		0.68		0.68		0.68	
Coef_G abs.	0.647		0.636		0.653		0.657		0.659		0.661	
rounded	0.65		0.64		0.65		0.66		0.66		0.66	
Rel. Err. Var.	0.097		0.102		0.095		0.093		0.092		0.092	
Rel. Std. Err. of M.	0.312		0.320		0.308		0.305		0.304		0.303	
Abs. Err. Var.	0.108		0.114		0.105		0.104		0.103		0.102	
Abs. Std. Err. of M.	0.329		0.337		0.325		0.322		0.320		0.319	

EK 7. Etik Kurul İzni

Evrak Tarih ve Sayısı: 30.12.2022-E.548756



T.C.
GAZİ ÜNİVERSİTESİ REKTÖRLÜĞÜ
Etik Komisyonu

Sayı : E-77082166-302.08.01-548756
Konu : Bilimsel ve Eğitim Amaçlı

30.12.2022

Dağıtım Yerlerine

Üniversitemiz Eğitim Bilimleri Enstitüsü Eğitim Bilimleri Ana Bilim Dalı **Yüksek Lisans Öğrencisi Serpil KOCAOĞLU'nun, Doç.Dr.Melek Gülşah ŞAHİN'in** danışmanlığında yürüttüğü **"Açık Uçlu ve Çoktan Seçmeli Maddelerden Oluşan Madde Takımlarının Genellenebilirlik Kuramı ile İncelenmesi"** adlı tez çalışması ile ilgili konu Komisyonumuzun **27.12.2022** tarih ve **22** sayılı toplantısında görüşülmüş olup,

İlgilinin çalışmasının, yapılması planlanan yerlerden izin alınması koşuluyla yapılmasında etik açıdan bir sakınca bulunmadığına oybirliği ile karar verilmiş ve karara ilişkin imza listesi ekte gönderilmiştir.

Bilgilerinizi rica ederim.



Araştırma Kod No: 2022 - 1510

Komisyon Başkanı

Ek:1 Liste

DAĞITIM

Gereği:

Sayın Doç. Dr. Melek Gülşah ŞAHİN

Bilgi:

Eğitim Bilimleri Enstitüsü Müdürlüğüne



T.C.
GAZİ ÜNİVERSİTESİ REKTÖRLÜĞÜ
Etik Komisyonu

Sayı : E-77082166-302.08.01-739338
Konu : Bilimsel ve Eğitim Amaçlı

07.09.2023

Dağıtım Yerlerine

Daha önce 30.12.2022 tarih ve E.548756 sayılı yazımız ile onay alan 2022-1510 kod numaralı, Üniversitemiz Eğitim Bilimleri Enstitüsü Eğitim Bilimleri Ana Bilim Dalı Yüksek Lisans Öğrencisi Serpil KOCAOĞLU'nun, Doç.Dr.Melek Gülşah ŞAHİN'in danışmanlığında yürüttüğü "*Açık Uçlu ve Çoktan Seçmeli Maddelerden Oluşan Madde Takımlarının Genellenabilirlik Kuramı ile İncelenmesi*" başlıklı tez çalışması hakkında ilgililerden alınan 31.08.2023 tarihli dilekçe Komisyonumuzun 05.09.2023 tarih ve 15 sayılı toplantısında görüşülmüş olup,

Dilekçede; tez savunma aşamasında tez başlığının "*Açık Uçlu ve Çoktan Seçmeli Maddelerden Oluşan Madde Takımlarının Güvenirliliğe Etkisinin Genellenebilirlik Kuramı ile İncelenmesi*" olarak değiştirilmesi ile ilgili talebin uygun olduğuna oybirliği ile karar verilmiş ve karara ilişkin imza listesi ekte gönderilmiştir.

Bilgilerinizi rica ederim.

Komisyon Başkanı

Ek: 1 Liste

DAĞITIM

Gereği:

Sayın Doç. Dr. Melek Gülşah ŞAHİN

Bilgi:

Eğitim Bilimleri Enstitüsü Müdürlüğüne

EK 8. MEB Araştırma İzni

Evrak Tarih ve Sayısı: 17.02.2023-E.590896



T.C.
ANKARA VALİLİĞİ
Milli Eğitim Müdürlüğü

Sayı : E-14588481-605.99-70491945

16.02.2023

Konu : Araştırma İzni

GAZİ ÜNİVERSİTESİ REKTÖRLÜĞÜNE
(Eğitim Bilimleri Enstitüsü Müdürlüğü)

İlgi: a) MEB Yenilik ve Eğitim Teknolojileri Genel Müdürlüğü'nün 2020/2 sayılı Genelgesi.
b) 13.02.2023 tarihli ve 585860 sayılı yazınız.

Üniversiteniz Eğitim Bilimleri Ana Bilim Dalı, Eğitimde Ölçme ve Değerlendirme Bilim Dalı Yüksek Lisans öğrencisi Serpil KOCAOĞLU'nun "**Açık Uçlu ve Çoktan Seçmeli Maddelerden Oluşan Madde Takımlarının Genellenebilirlik Kuramı İle İncelenmesi**" konulu çalışması kapsamında Beypazarı ilçemize bağlı resmi ortaokullarda uygulama yapma talebi ilgi (a) Genelge çerçevesinde incelenmiştir.

Yapılan inceleme sonucunda, söz konusu araştırmanın Müdürlüğümüzde muhafaza edilen ölçme araçlarının; Türkiye Cumhuriyeti Anayasası, Milli Eğitim Temel Kanunu ile Türk Milli Eğitiminin genel amaçlarına uygun olarak, ilgili yasal düzenlemelerde belirtilen ilke, esas ve amaçlara aykırılık teşkil etmeyecek, eğitim-öğretim faaliyetlerini aksatmayacak şekilde okul ve kurum yöneticilerinin sorumluluğunda, gönüllülük esasına göre uygulanması Müdürlüğümüzce uygun görülmüştür.

Bilgilerinizi ve gereğini rica ederim.

Harun FATSA
Vali a.
Milli Eğitim Müdürü

Ek: Uygulama Araçları (18 Sayfa)

Dağıtım:
Gereği:
Gazi Üniversitesi

Bilgi:
Beypazarı İlçe MEM



GAZİLİ OLMAK AYRICALIKTIR