



**T.C.**  
**GAZİOSMANPAŞA ÜNİVERSİTESİ**  
**EĞİTİM BİLİMLERİ ENSTİTÜSÜ**  
**EĞİTİM BİLİMLERİ ANABİLİM DALI**  
**EĞİTİMDE ÖLÇME VE DEĞERLENDİRME YÜKSEK LİSANS**  
**PROGRAMI**

**ÖĞRENCİ BAŞARISININ BELİRLENMESİNDE**  
**KULLANILAN FARKLI ÖLÇME ARAÇLARINDAN ELDE**  
**EDİLEN PUANLARIN KARŞILAŞTIRILMASI**

**YÜKSEK LİSANS TEZİ**

**Muammer YÜKSEL**

**TOKAT**  
**Mayıs, 2015**



**T.C.**  
**GAZİOSMANPAŞA ÜNİVERSİTESİ**  
**EĞİTİM BİLİMLERİ ENSTİTÜSÜ**  
**EĞİTİM BİLİMLERİ ANABİLİM DALI**  
**EĞİTİMDE ÖLÇME VE DEĞERLENDİRME YÜKSEK LİSANS**  
**PROGRAMI**

**ÖĞRENCİ BAŞARISININ BELİRLENMESİNDE**  
**KULLANILAN FARKLI ÖLÇME ARAÇLARINDAN ELDE**  
**EDİLEN PUANLARIN KARŞILAŞTIRILMASI**

**YÜKSEK LİSANS TEZİ**

**Muammer YÜKSEL**

**Danışman: Doç. Dr. Gülşah BAŞOL**

**TOKAT**  
**Mayıs, 2015**

## JÜRİ ÜYELERİNİN İMZA SAYFASI

Eğitim Bilimleri Enstitüsü Müdürlüğüne,

Muammer YÜKSEL 'in ' Öğrenci Başarısının Belirlenmesinde Kullanılan Farklı Ölçme Araçlarından Elde Edilen Puanların Karşılaştırılması' adlı çalışması 26/05/2015 tarihinde jürimiz tarafından Eğitim Bilimleri Anabilim Dalı'nda yüksek lisans tezi olarak kabul edilmiştir.

Adı Soyadı

İmza

Başkan: Prof. Dr. Mehmet ARSLAN

Üye (Tez Danışmanı): Doç Dr. Gülşah BAŞOL

Üye : Yrd. Doç. Dr. Hacı Bayram YILMAZ

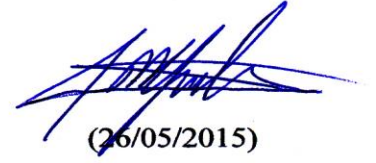
Onay

Yukarıdaki imzaların adı geçen öğretim üyelerine ait olduğunu onaylarım.

06/05/2015  
Yrd. Doç. Dr. Yasir. GÖRÜŞÜM  
Enstitü Müdürü

## ETİK SÖZLEŞME

Bu belge ile bu tezdeki bütün bilgi toplama ve raporlaştırma sürecinin Gaziosmanpaşa Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğine, Eğitim Bilimleri Enstitüsü Tez Yazım Kılavuzuna, genel akademik kurallara ve etik ilkelere uygun olarak gerçekleştirildiğini; bu tez çalışmasını “intihali engelleme” programı ile taradığımı, bana ait olmayan tüm bilgi, düşünce ve bulgulara atıf yaptığımı ve kaynağını gösterdiğimi beyan eder, sorumluluğun tarafıma ait olduğunu kabul ederim.



(26/05/2015)

Muammer YÜKSEL

**Her zaman yanımda olduklarını bildiklerime**

**AİLEME...**

## TEŞEKKÜRLER

Çalışma azmi, desteği ve önerileriyle çalışmamı tamamlamama katkıda bulunan, öğrencilerine verdiği önemle ve kişiliğiyle takdir ettiğim, araştırmanın analizleri sürecinde yardımını esirgemeyen, bilgilerini her daim paylaşan ve madde analizlerini yapmama yardımcı olan, EduG, TAP.exe programlarına erişimimi sağlayıp kullanımlarını öğreten, verdiği dönütlere çalışmama katkılarıyla nitelikli bir bilimsel çalışmanın nasıl olması gerektiğini bana öğreten ve benimseten, İngilizce kaynak tarama ve çeviri sürecinde yardımlarını esirgemeyen; danışmanım ve hocam Sayın Doç. Dr. Gülşah BAŞOL' a;

Araştırma verilerinin analizleri için gerekli EduG programının kullanımı sürecinde destek veren, analizlerin yapım aşamasında yardımını esirgemeyen ve yine yapmış olduğu alanyazın çalışmaları ile çalışmama destek ve yardımcı olan Doç. Dr. Neşe GÜLER' e;

Eğitim hayatım boyunca desteğini her zaman yanımda hissettiğim ve benden hiçbir zaman yardımını esirgemeyen biricik ABLAM' a;

Varlığıyla eksik yanımlı tamamlayan, hayatımın mutlu ve huzurlu olmasına katkı sağlayan biricik EŞİM' e;

Bana her konuda güvenen, hep daha ileriye gitmem için beni cesaretlendiren, üzerimden desteğini ve yardımlarını esirgemeyen ANNEM ve BABAM' a;

Madde seçim sürecinde ve konu alanı taramasında yardımcı olan sevgili Ayda GÖÇMEN' e;

Ölçme ve Değerlendirme bölümünü bana sevdiren ve alana yönelmemi sağlayan, her zaman derslerdeki neşesi ile örnek aldığım hocam Sayın Yrd. Doç. Dr. Ümit ÇELEN' e ve Araş. Gör. Ayşe DÖNMEZ' e, düzeltmelerde yardımcı olan Türkçe Öğretmeni Sayın Seyhan GÜNAYDIN' a;

Buraya ismini sığdıramadığım desteğini, sevgisini ve bilgisini benimle paylaşan tüm hocalarıma ve arkadaşlarıma;

SONSUZ TEŞEKKÜRLERİMİ SUNARIM.

## ÖZET

### ÖĞRENCİ BAŞARISININ BELİRLENMESİNDE KULLANILAN FARKLI ÖLÇME ARAÇLARINDAN ELDE EDİLEN PUANLARIN KARŞILAŞTIRILMASI

Yüksel, Muammer

Yüksek Lisans, Eğitimde Ölçme ve Değerlendirme Bilim Dalı

Tez Danışmanı: Doç. Dr. Gülşah Başol

Mayıs 2015 , ix + 85 sayfa

Araştırmanın temel amacı; genellenebilirlik kuramına göre öğrencilerin performanslarının ölçülmesinde kullanılan farklı ölçme araçlarının G ve K çalışmaları sonuçlarını karşılaştırmaktır.

Araştırmanın çalışma grubunu, 2013- 2014 öğretim yılı Kars il merkezinde bulunan Atatürk Ortaokulu'nda öğrenim gören sekizinci sınıf öğrencilerinden rastgele seçilmiş 48 öğrenci oluşturmaktadır. Öğrencilere araştırmacı tarafından geliştirilen çoktan seçmeli test ve klasik sınavın uygulandığı araştırmada, ölçme araçlarının puanlanmasında üç puanlayıcı görev almıştır. Klasik sınavı üç farklı puanlayıcı puanlamış ve ölçme sonuçları üzerinde G ve K çalışması yapılmıştır.

Elde edilen sonuçlar, uygulanan çoktan seçmeli ve klasik sınav için karışan hata varyanslarının klasik sınavda daha çok olduğunu, puanlama ortalamasına bakıldığında ise çoktan seçmeli teste ait puan değerlerinin daha yüksek olduğunu ve varyans değerlerinin iki sınav için de paralellik gösterdiğini ortaya koymuştur.

Araştırmada ayrıca klasik sınav için farklı puanlayıcıların puanları arasındaki fark tespit edilmeye çalışılmıştır. Bunun yanı sıra karar çalışması yapılarak her iki ölçme aracı için bulunan G ve Phi katsayılarındaki farklılıklar incelenmiştir. Buna göre, her iki ölçme aracı için sonuçlar paralellik göstermektedir. Çoktan seçmeli test ve klasik sınav için yapılan karar çalışması sonuçlarına göre madde sayısının artırılması halinde Phi ve G katsayılarının arttığı; yapılan genellenebilirlik çalışmasına göre her iki sınav türü için de ana varyans kaynağı açısından öncelikli etkinin bireylerden kaynaklandığı görülmüştür. Ayrıca araştırmada farklı puanlayıcılarla tutarlı sonuçlar elde edildiği görülmüştür. Buna göre, soru sayısının arttırılmasındansa puanlayıcı sayısının arttırılmasıyla elde edilen güvenilirliğin klasik sınav için daha yüksek olacaktır. Soru sayısını artırılması yazılı sınavların okunmasını güçleştirdiğinden imkanları zorlayarak puanlayıcı sayısının arttırılması önerilmiştir.

**Anahtar Kelimeler:** Genellenebilirlik Kuramı, Klasik Sınav, Başarı.

## **ABSTRACT**

### **COMPARISON OF SCORES OBTAINED FROM DIFFERENT MEASUREMENT TOOLS USED IN THE DETERMINATION OF STUDENT ACHIEVEMENT**

Yüksel, Muammer

Graduate Degree, Measurement and Evaluation in Education

Thesis Advisor: Doç. Dr. Gülşah Başol

May 2015 , ix + 85 sayfa

This is a generalizability study, measuring students' performance through different measurement tools and comparing the finding through G theory and K in order to identify the errors associated with the raters and items to improve the future applications.

The sample consisted of 48 randomly selected 8th graders, registered to Atatürk Secondary School, located in the center of Kars province during 2013-2014 academic year. Two different types of exams (a multiple choice test and an essay) prepared by the researcher were applied to students. Three raters graded the students' exams. Essays were scored by three different raters and G and K analyses were performed on the results.

According to the findings, the error rate was higher for the essays in comparison to multiple choice test. The mean score was higher for the multiple choice test, on the other hand the variances were found to be similar.

For the essays, there were no differences among the scores given by different coders. Decision studies indicated that the findings were in line for both type of the measurements according to G and Phi coefficients. The results indicated that increasing the number of item were resulted in an increase in the G and Phi coefficients. According to the findings, the Person Factor was the main source of the variation in the data for both measurement type. As the findings indicated consistent result among the coders, we suggest to increase the number of coders rather than increasing the number of questions for the essay. Considering the excessive work, associated with grading an essay with more question, when available, it can be better to increase the number of raters.

**Key Words:** Generalizability Theory, Essay, Success.



## İÇİNDEKİLER

ETİK SÖZLEŞME .....	I
TEŞEKKÜR .....	III
ÖZET ERROR! BOOKMARK NOT DEFINED.....	
.IV	
ABSTRACT.....	V
İÇİNDEKİLER .....	VI
TABLolar LİSTESİ .....	IX
<b>1. GİRİŞ.....</b>	<b>1</b>
1.1. Problem Durumu.....	1
1.2. Araştırmanın Amacı.....	5
1.3. Araştırmanın Önemi.....	5
1.4. Sayıtlar.....	6
1.5. Sınırlılıklar.....	6
1.6. Tanımlar.....	6
<b>2. KAVRAMSAL ÇERÇEVE.....</b>	<b>8</b>
2.1. Ölçme ve Değerlendirme.....	8
2.2. Ölçme Araçları.....	11
2.2.1. Klasik ölçme araçları .....	12
2.2.1.1. Yazılı yoklamalar .....	12
2.2.1.2. Çoktan seçmeli sınavlar.....	14
2.2.1.3. Doğru yanlış testleri .....	16
2.2.1.4. Eşleştirme testleri .....	16
2.2.1.5. Kısa cevaplı testler.....	17
2.2.1.6. Matematikte problem çözme testleri .....	17
2.2.2. Alternatif ölçme araçları .....	17
2.2.2.1. Tanılayıcı dallanmış ağaç.....	19
2.2.2.2. Yapılandırılmış grid.....	20
2.2.2.3. Kelime ilişkilendirme .....	20
2.2.2.4. Portfolyo (Öğrenci ürün dosyası).....	21

2.2.2.5. Proje.....	22
2.2.2.6. Performans deęerlendirme .....	23
2.2.2.7. Gzlem .....	23
2.2.2.8. Kavram haritası .....	24
2.2.2.9. Grüşme .....	25
2.3. Genellenebilirlik kuramı .....	26
2.3.1. Facet.....	28
2.3.2. aprazlanmış ve Yuvalanmış Desen .....	28
2.3.3. Tesadüfi ve Sabit Deęişkenlik Kaynakları.....	29
2.3.4. G alışması ve K alışması .....	29
2.4. İlgili Araştırmalar .....	33
2.4.1. Genellenebilirlik Kuramı İle İlgili Araştırmalar .....	33
2.4.2. Farklı Ölçme Araçları İle İlgili Araştırmalar .....	41
<b>3. YÖNTEM .....</b>	<b>44</b>
3.1. Araştırmanın Modeli .....	44
3.2. Evren ve Örneklem .....	44
3.3. Araştırma Verileri .....	45
3.4. Verilerin Analizi .....	48
<b>4. BULGULAR .....</b>	<b>49</b>
4.1. Alt Problem 1; Performansın Ölçülmesinde Kullanılan Çoktan Seçmeli Sınavın İncelenmesi .....	49
4.1.1. Performansın Ölçülmesine Yönelik G Kuramına Göre Kestirilen Parametrelerin Varyansları ve Toplam Varyansları Açıklama Yüzdeleri.....	50
4.1.2. Performansın Ölçülmesine Yönelik Madde Sayılarının Arttırılıp Azaltılması Senaryolarına Göre Hesaplanan Hata Varyansları.....	51
4.1.3. Performansın Ölçülmesine Yönelik Madde Sayılarının Arttırılıp Azaltılması Senaryolarına Göre Hesaplanan G ve $\Phi$ Katsayıları .....	52
4.2. Alt Problem 1; Performansın Ölçülmesinde Kullanılan Klasik Sınavın İncelenmesi .....	53

4.2.1. Performansın Ölçülmesine Yönelik G Kuramına Göre Kestirilen Parametrelerin Varyansları ve Toplam Varyansları Açıklama Yüzdeleri .....	55
4.2.2. Performansın Ölçülmesine Yönelik Madde Sayılarının Arttırılıp Azaltılması Senaryolarına Göre Hesaplanan G ve $\Phi$ Katsayıları .....	57
4.2.3. Performansın Ölçülmesine Yönelik Madde sayılarının Arttırılıp Azaltılması Senaryolarına Göre Hesaplanan Mutlak ve Bağıl Hata Varyansları .....	59
<b>5. TARTIŞMA.....</b>	<b>61</b>
<b>6. SONUÇLAR VE ÖNERİLER.....</b>	<b>65</b>
6.1. Sonuçlar .....	65
6.1.1. Alt Probleme Ait Sonuçlar .....	65
6.1.2. Alt Probleme Ait Sonuçlar .....	65
6.1.3. Alt Probleme Ait Sonuçlar .....	66
6.2.1. Alt Probleme Ait Sonuçlar .....	66
6.2.2. Alt Probleme Ait Sonuçlar .....	67
6.2.3. Alt Probleme Ait Sonuçlar .....	68
6.2. Öneriler .....	69
6. 2. 1. Araştırmadan Çıkan Öneriler .....	69
6. 2. 2. İleride Yapılacak Araştırmalar İçin Öneriler .....	69
<b>KAYNAKÇA.....</b>	<b>70</b>
<b>EKLER.....</b>	<b>74</b>
<b>ÖZGEÇMİŞ.....</b>	<b>85</b>

## TABLolar LİSTESİ

Tablo 1. Deneme Uygulamasından Elde Edilen Madde İstatistikleri .....	45
Tablo 2. Çoktan Seçmeli Testin Ön Uygulamasına ait Betimsel İstatistikler .....	47
Tablo 3. Çoktan Seçmeli Test İle Yapılan Nihai Uygulamaya Ait Betimsel İstatistikler .....	49
Tablo 4. Tek Değişkenli G Çalışması Sonucunda Ölçmenin Kestirilen Varyansları ve Toplam Varyansı Açıklama Oranları .....	50
Tablo 5. Performansın Ölçülmesine İlişkin Yapılan K Çalışması İle Madde Sayıları Senaryolarına Göre Bağıl ve Mutlak Hata Varyansları .....	51
Tablo 6. Performansın Ölçülmesine İlişkin Yapılan K çalışması İle Madde Sayıları Senaryolarına Göre G ve Phi Katsayıları.....	53
Tablo 7. Performansın Ölçülmesinde Yapılan Klasik Sınav İçin 3 Puanlayıcı Ait Betimsel İstatistikler .....	54
Tablo 8. Puanlayıcıların 11 Maddeye Verdikleri Puanlar Arasındaki Korelasyon Katsayıları .....	55
Tablo 9. Tek Değişkenli G Çalışması Sonucunda Ölçmenin Kestirilen Varyansları ve Toplam Varyansı Açıklama Oranları .....	55
Tablo 10. Performansın Ölçülmesine İlişkin Yapılan K çalışması ile Madde ve Puanlayıcı Sayıları Senaryolarına Göre Phi ve G Katsayıları .....	57
Tablo 11. Performansın Ölçülmesine İlişkin Yapılan K Çalışması İle Madde ve Puanlayıcı Sayıları Senaryolarına Göre Bağıl ve Mutlak Hata Varyansları .....	59

## **BÖLÜM I GİRİŞ**

Bu bölümde, araştırmanın problem durumuna, problem cümlesine ve araştırmadaki alt problemlere, araştırmanın amacı ve önemine, sayıtlılara, sınırlılıklara, kısaltmalara ve ilgili araştırmalara yer verilmiştir.

### **1. 1. Problem Durumu**

Eğitim bireylerin hayatları boyunca devam eden bir süreçtir. Gerek formal gerekse informal olarak gerçekleşebilen eğitim bireyin gelişim sürecinin ayrılmaz bir parçasıdır. Öğrenme, eğitim süreci sonucunda ve belli yaşanmışlıkların etkisi ile ortaya çıkan, bireye yeni nitelik kazandırma sürecine denir.

Son yıllarda ön plana daha çok çıkan yaşam boyu öğrenme anlayışı, bireylerin gelişmelerine büyük oranda katkı sağlamaktadır. Hayatlarının her safhasında bireylere yeni şeyler öğrenme ve kendini geliştirme imkanı sağlayan bu görüş, öğrenmelerin ve eğitimin önemini daha çok ortaya çıkarmaktadır. Eğitim süreci bünyesinde yer alan hazırbulunuşluk, güdülenme, öğretim, ölçme ve değerlendirme gibi kavramlar da artık hayatın içinde daha sık benimsenmektedir. Her biri birbirinin tamamlayıcısı ve önemli bir parçası olan içi içe geçmiş bu süreçleri bilmek eğitim ve öğretimin kalitesini artırır.

Eğitim istendik davranış oluşturma veya istendik davranış değiştirme süreci olarak, toplumun süzgeçten geçirilmiş değerlerinin, ahlak standartlarının, bilgi ve beceri birikimlerinin yeni nesillere aktarılmasıdır (Senemoğlu, 2002). Eğitim süreci sonunda bireylerin belli konularda bilgi, beceri ve tutum kazanması beklenir. Eğitim sürecindeki bireylerden kazanılması beklenen bu bilgi, beceri ve tutumların kazanılma düzeyinin; sürecin verimliliğini göstermesi ve dönüt sağlayarak süreci zenginleştirilmesi beklenmektedir. Eğitim sistemimizde bireylerin bu kazanımları başarı olarak nitelendirilmekte ve başarı düzeyleri belirlenmeye çalışılmaktadır.

Etkinlikler sonunda beklenen kazanımların; bir kısmının oluştuğu, bir kısmının yeterli düzeyde oluşmadığı, istenmeyen kazanım şeklinde ortaya çıktığı veya planlandığı şekilde oluşmadığı görülmektedir. Bu durum eğitimde kontrol ihtiyacını doğurur (Turgut ve Baykul, 2010). Burada yer alan kontrol kavramı eğitim sürecinin ve

ürünlerinin gözden geçirilmesi ve bir sonuca varılması anlamına gelmektedir denilebilir. Kontrol süreci eğitimi hem planlı hale getirir hem de var olan eksikliklerin giderilmesine ve kalitenin artırılmasına olanak sağlar.

Öğrencilerin başarılarının belirlenmesinde öncelikle ölçme ve sonrasında bunu da içine alan değerlendirme sürecine yer verilmelidir. Eğitim sürecindeki bireylerin eğitimden ne kadar yararlandıkları ya da öğrenilmesi beklenen kazanımlara ne kadar ulaşıldığı sürekli merak konusu olmaktadır. Çünkü hem eğitimin niteliği hem de bireyler hakkında verilecek kararlar için kazanımların ulaşılma düzeyleri saptanmak istenmektedir. Burada da devreye ölçme ve değerlendirme süreci girer.

Kazanımla ifade edilen hedefleri gerçekleştirme yolunda öğretim etkinlikleri planlanır. Öğretimde izlenen yöntemi de dikkate alarak farklı ölçme araçları arasından, öğrenmenin gerçekleşip gerçekleşmediğini yoklamak için en uygun olanı seçilir. Değerlendirmenin amacına göre kullanılan ölçme araçları da çeşitlilik gösterir. Ölçme yönteminin hedeflenen kazanımlara uygun olması ölçüm sonuçlarının geçerliliği için önemlidir. Bu nedenle ölçülmek istenilen kazanımların niteliğine en uygun olabilecek ölçme aracının seçilmesine gerekli önem verilmelidir.

Kullanılacak ölçme aracına öğrencinin hazırbulunuşluk düzeyi, sınavın yapılacağı ortam, zaman sınırlaması olup olmadığı ve uygulama koşulları gibi faktörler dikkate alınarak karar verilir. Farklı ölçme araçlarıyla elde edilen sonuçların benzer olup olmadığı araştırılmaya değer bir konudur. Bu sayede farklı kaynaklardan ulaşılan ölçüm sonuçlarının güvenilir olup olmadığını anlamak mümkün olur.

Eğitim sürecinin önemli bir parçası olan ölçme için bireylerin farklı ölçmeler sonucunda ortaya çıkan sonuçlarının birbirleri ile ilişkisinin nasıl olduğu bir merak konusudur. Bu soruların cevaplarının bulunması elbette performans, not ve başarı seviyesi olarak ilerleyen sürecin daha anlamlı şekilde açıklanmasını sağlayabilir.

Katılımcıların performansının ölçülmesinin amaçlandığı araştırmalarda genellikle araştırmacılar Klasik Test Kuramı (KTK) kullanarak istatistiksel çalışmalar yapmışlardır. KTK araştırmacılar tarafından daha çok bilinmesi ve temel alınması sebebi ile tercih edilirken bazı sınırlılıkları araştırmanın uzamasına ya da istenilen

sonulara tam olarak yaklařılamamasına neden olmaktadır. Genellenebilirlik Kuramı (G Kuramı) ise hata kaynaklarını aynı anda ele alması ve birbirleri ile iliřkilerine yer vermesi nedeni ile arařtırma sonuları iin bir karřılařtırmaya olanak vermektedir. Ayrıca G Kuramı ile arařtırma sonularıyla sonular üzerinden farklı senaryolar üretilebilir. Üretilen bu senaryolardan yola çıkılarak farklı sayıdaki deęiřken miktarları ve farklı arařtırma durumlarında parametrelerin deęerleri kestirilebilir. Böylece yapılan bir arařtırmadan yola çıkılarak gelecekte yapılabilecek arařtırmalar iin sonular kestirilerek en doęru seim yapılabilir ve arařtırma yolunun belirlenmesi saęlanmış olur.

Cronbach ve arkadaşları tarafından 1963 yılında temelleri atılan G Kuramı ile ilgili alıřmalar yurt dıřında aynı tarihleri takriben bařlar iken ölkemizde 2004 yılından itibaren ve daha ok yüksek lisans, doktora tezleri üzerinde yoğunluk göstermiřtir. Bu yeni kuram; bařlarda tezlerde yapılan arařtırmalarla, günümüzde makalelerle ve üzerine yazılan bir kitap ile (Güler, Kaya Uyanık, Tařdelen Teker, 2012) arařtırmacılara kendini benimsetmiřtir.

LLabre 1978’de yapmıř olduęu G Kuramı yardımıyla analizlerini gerekleřtirdięi alıřmasında; öęrencilerin yazma becerilerini etkileyen deęiřkenlerin özelliklerini belirlemek iin farklı özelliklerdeki sınıflarda bulunan öęrencileri guruplara ayırmıř ve üç farklı mevsimde, farklı özellikteki yazma konuları üzerine verdikleri cevapları 4 farklı puanlayıcı tarafından deęerlendirilmesini saęlayarak deęiřkenlerin birbirleri ile ve ayrı ayrı varyans deęerlerini hesaplamıřtır. Hesaplanan bu deęerlerden yola çıkarak farklı senaryolara göre hangi durumlarda öęrencilerin yazma becerilerinin artış göstereceęini kestirmeye alıřmıřtır.

Ölkemizde daha yeni yeni yangınlařmaya bařlayan G Kuramı alıřmaları genellikle performansın ölçölme süreci, puanlayıcılar ve klasik ölçme araçları üzerinde yoğunlařmıřtır. Puanlayıcıların, bireyler ve maddelerin etkileri arařtırılır iken farklı desenlerin incelendięi arařtırmalar ( Wang, 2005; Au, Prahardhi ve Shiell 2008; Lane ve Sabers, 1989; Nalbantoęlu Yılmaz ve Uzun Bařusta, 2012; Nalbantoęlu , 2009) daha ok yoğunluk kazanmaktadır.

Atılgan, 2004; Güler, 2008 ve Alkahtani 2012' de G Kuramı ile yaptıkları çalışmalarında arařtırmacılar KTK yanında Çok Deęişkenli Rasch Modeli (ÇDRM) kullanmışlar; maddelerin zorluk düzeyleri ve puanlayıcıların puan verme eğilimleri hakkında bilgiye ulaşmaya çalışmışlardır.

Kuramların ve modellerin karşılaştırılmasının yanında, bazı çalışmalarda Lojistik Regresyon Analizi kullanılması, farklı kesme puanları hesaplama yöntemlerinin karşılaştırılması, farklı ölçeklerin güvenilirliklerinin araştırılması çalışmaları G Kuramı yardımıyla yapılmıştır.

Lynch ve McNarama (1998), Molloy and Shimura (2005), Lee (2005), Sawaki (2007) ve Kim (2009) farklı yıllarda dil testi ile yaptıkları konuşma becerisi üzerine olan arařtırmalarında madde, görev, değerlendirici gibi farklı hata kaynaklarını Genellenebilirlik Kuramı ile değerlendirmişlerdir. Özellikle, değerlendirici hata kaynağı üzerinde bütün arařtırmalarda durulmuştur. Lin ve Zhang (2014) Almanca öğreniminde 20 çalışmada 500 değerlendirmecinin yaptıkları değerlendirmelerin güvenilirliğini ve ölçümlerin standart hatasını genellenebilirlik analizi yaparak arařtırmışlardır. Sonuç olarak, 3-6 değerlendirmeci ile kabul edilebilir düzeyde bir güvenirligi yakalamanın mümkün olacağını ve ölçümlerin standart hatasının da bu şekilde aşağıya çekileceğini belirtmişlerdir.

Huang (2012) genellenebilirlik kuramına göre geniş ölçekli bir Yabancı Dil Sınavının (ESL) güvenirligi ve geçerliğinin arařtırdığı çalışmasının sonucu olarak, ilgili sınavın geçerliği konusunda problemleri ortaya koyarak potansiyel bir yanılgıya işaret etmiştir. Genellenebilirlik çalışmaları sayesinde testlerin daha etkili ölçümler yapabilecek şekilde dizayn edilmeleri mümkündür. Örneğin, Neill ve Neill (2015)' in çalışmasında tek bir test bataryası kullanmak yerine kompozite puan veren bir test bataryası kullanıldığında ölçümlerin güvenirliginin artacağı ve etkili sonuçlar alınacağı belirtilmiştir. Bottema-Beutel, Lloyd, Carter, ve Asmus (2014) Genellenebilirlik Kuramını gözlemsel ölçümlerin güvenirliginin değerlendirmesinde kullanmışlar ve pilot çalışmadan elde edilen sonuçlardan yapılan güvenirlilik tahminlerinin önemini altını çizmişlerdir.



Farklı ölçme araçlarının aynı konu üzerindeki etkilerinin araştırıldığı çalışmalara bakıldığında; çalışmaların daha çok kısa cevaplı, çoktan seçmeli, eşleştirmeli, klasik (yazılı) sınavlar ve performans ödevleri üzerinde arttığı görülmüştür. Demir' in 2010 yılında yaptığı çalışmasında; uluslar arası sınavlara ait verilerin kullanıldığı ve farklı soru tiplerinin, öğrencilerin verdikleri cevaplara göre başarı durumları incelenmiştir. Yapılan bu araştırmalarda hata kaynakları KTK' ya göre ele alındığı için her bir hata kaynağı için ayrı hata varyansları hesaplanmış ancak ortak hata varyans değerleri hesaplanmamıştır.

Öğrencilerin aynı konu içeriğine sahip farklı özellikteki ölçme araçları ile başarıları belirlenmek istense acaba sonuçlar aynı olur mu? Farklı kaynaklardan elde edilen puanlar birbiriyle tutarlı mıdır? Karşılaştırma farklı kuramlar çerçevesinde yapılırsa parametrelerin değişimi acaba nasıl olur?

Mevcut araştırmanın amacı, aynı konu içeriğine sahip farklı özellikteki ölçme araçlarından elde edilen başarı puanlarının birbirleri ile tutarlılığını araştırmaktır. Ayrıca farklı özellikteki ölçme araçları ile aynı bireyler ve durumlarda elde edilen ölçme sonuçlarına ait hata miktarının her bir değişken için ayrı ayrı ve birbirleri ile kombinasyonlarının belirlenmesi amaçlanmıştır. Farklı ölçme araçları ve birden fazla hata kaynağı ele alınacağı için bu araştırma Genellenebilirlik Kuramı temel alınarak gerçekleştirilmiştir.

## **1. 2. Araştırmanın Amacı**

Bu araştırmanın amacı; öğrencilerin başarılarının belirlenmesinde kullanılan farklı iki ölçme tekniğinden hangilerinin daha güvenilir sonuçlar vereceğinin Genellenebilirlik Kuramı kullanılarak ortaya çıkarılmasıdır.

### **Alt Problemler**

1.1. Çoktan seçmeli test için G Kuramına göre kestirilen parametrelerin varyansları ve toplam varyansları açıklama yüzdeleri nedir?

1.2. Çoktan seçmeli test için yapılan K çalışması doğrultusunda maddelerin hata varyanslarının değişimleri nasıldır?

1.3.Çoktan seçmeli test için yapılan K çalışması sonuçlarına göre G ve Phi katsayılarının değişimleri nasıldır?

2.1. Klasik sınav için yapılan G Kuramına göre kestirilen parametrelerin varyansları ve toplam varyansları açıklama yüzdeleri nedir?

2.2. Klasik sınav için yapılan K çalışmasına göre farklı senaryolara göre G ve Phi katsayılarının değişimi nasıldır?

2.3. Klasik sınav için yapılan K çalışması doğrultusunda maddelerin hata varyanslarının değişimleri nasıldır?

### **1. 3. Araştırmanın Önemi**

Bu araştırmayla farklı niteliğe sahip ölçme araçları ile elde edilen sonuçların tutarlılığının G Kuramı analizleri ile ortaya çıkarılabileceği düşünülmektedir. G Kuramı güvenilirlik bağlamında daha ayrıntılı bilgiler verdiği için bu çalışmada kullanılmıştır. Ayrıca birbirinden farklı ölçme araçları puanlarına karışan hata puanlarının (Çoktan seçmeli test için madde, bireyden; klasik sınav için madde, birey ve puanlayıcıdan) farklılık gösterip göstermediğinin ortaya çıkarılabileceği düşünülmektedir. Ölçme araçlarına karışan hata miktarlarına bakılarak ölçme araçlarından hangisinin daha güvenilir olarak kullanılabileceği ve sınav türlerine göre hangi değerlendirme türünün kullanımının daha doğru olabileceği hakkında uygulamalara ışık tutması bakımından önem taşımaktadır.

### **1. 4. Sayıtlar**

Katılımcılar uygulamada cevaplandırmış oldukları soruları, yeterli motivasyon ve özveri ile yanıtlamışlardır.

Öğrencilerin cevap kağıtlarından aldıkları puanlar öğrencilerin konu hakkındaki bilgi düzeylerini yansıtmaktadır.

Puanlayıcılar öğrencileri objektif olarak puanlamışlardır.

### **1. 5. Sınırlılıklar**

Araştırma 2013-2014 eğitim- öğretim yılı Kars ili Atatürk Ortaokulu 8. sınıf öğrencilerinden seçilen 48 kişi ile ve Fen ve Teknoloji dersi 8. sınıf ‘ *Hücre Bölünmesi ve Kalıtım* ’ ünitesi ile sınırlıdır.

## 1. 6. Tanımlar

**Genellenebilirlik Kuramı:** Genellenebilirlik (G) Kuramı ölçme sonuçlarının güvenilirliğinin belirlenmesini, güvenilir gözlemlerin tasarımını, araştırılmasını ve kavramsallaştırılmasını sağlayan istatistiksel bir kuramdır (Güler, 2008).

**G Katsayısı:** G katsayısı, bir bireyin gözlenen puanından o bireyin evren puanının ne derece doğru genellendiğinin bir göstergesidir. Klasik test kuramındaki güvenilirlik katsayısı gibi, genellenebilirlik katsayısı bireylerin puanlarındaki çeşitliliğin oranını yansıtır (Güler, 2008). Genellenebilirlik katsayısı (G-katsayısı), klasik test kuramındaki gerçek varyansın gözlenen varyansa oranı olan, güvenilirlik katsayısıyla benzerlik gösterir.

G katsayısının tanımı ölçmenin nasıl kullanılacağına bağlıdır. Hata varyansı göreceli ve mutlak modellerde farklılık gösterdiği için, G katsayısının büyüklüğü de kullanılan modele bağlı olarak değişir (Shavelson ve Webb, 1991; 14). Genellenebilirlik katsayısı, evren puanının varyansının gözlenen puanların varyansına oranlanması olarak tanımlanabilir.

**Phi- katsayısı:** Mutlak modeller için kullanılmakta olan ve gerçek varyansın, gerçek varyans ile mutlak varyansın toplamına bölünmesiyle elde edilen güvenilirlik indeksine denir.

**Ölçme:** Bir niteliğin gözlenip, gözlem sonucunun sayı ve sembollerle gösterilmesidir (Turgut, 1977).

## BÖLÜM II

### KAVRAMSAL ÇERÇEVE

#### 2. 1. Ölçme ve Değerlendirme

Ağırlık, uzunluk ve sayılabilecek fiziksel çokluklar gözlemlenerek sayılarla ya da sembollerle gösterilir. Ancak başarı, zeka, tutum ve yetenek gibi bireylerin psikolojik durumları doğrudan gözlenmez. Buna göre fiziksel ölçmelerde ölçmeye konu olan nitelikler doğrudan gözlenebilmesine karşın psikolojik ölçmelerdeki nitelikler dolaylı olarak gözlenip, gözlem sonuçları sayı ve sembollerle ifade edilir (Aiken, 1991).

Ölçmenin diğer gözlem tekniklerinden üstün olmasının nedeni, daha güvenilir, objektif ve kesin olmasıdır. Bu da, elde edilen verilerin sayı türünden ifade edilebilmesinden kaynaklanmaktadır. Toplanan verilerin sayı türünden ifade edilebilmesi ise, istatistiksel çözümlenme ve yorumlamayı kolaylaştırır, yargılarda daha kesin ve açık olmayı sağlar (Mert, 2008).

Her ölçme işleminde bir tür ölçek kullanılır. Ölçek, bir rakam veya sembol sistemidir. Bu sınıflamanın temel kriteri, ölçülen özelliğe ait ölçümlerin matematiksel özelliğidir (Çelik, 2000). Kullanım alanlarına göre ya da niteliğin alabileceği durumlara göre farklı ölçek türleri bulunmaktadır. Sosyal bilimlerde genel olarak kullanılan ölçek türleri şunlardır:

1. Oran ölçeği
2. Eşit aralıklı ölçek
3. Sıralama ölçeği
4. Sınıflama ölçeği.

Sözü edilen ölçek türlerinden eğitim alanında en çok ilk ikisi kullanılmaktadır. Eğitimde gerçekleştirilmesine çalışılan davranış değişiklikleri genellikle nicel yönleri baskın olan ya da en azından, nicel olarak betimlenebilen değişikliklerdir (Özçelik, 1998). Bu yüzden eğitimde en çok kullanılan sonuçlar, eşit aralıklı ve oran ölçeklerinden elde edilen ölçümlerdir.

Eđitimde sıklıkla kullanılan eşit aralıklı ölçek düzeyindeki ölçümler hakkında küçük bir hususa dikkat çekilmesinde fayda vardır. Eşit aralıklı ölçekte öğrencileri verilen notların kıyaslamada notların birbiri ile kat ilişkisi ve *sıfır* kavramının işlevsel olmadığı unutulmamalıdır. Bazı yanlış değerlendirmelerin kaynağı ölçekten elde edilen *sıfır* puanından kaynaklanır. Bu ölçek düzeyinde *sıfır* kavramı asla bulunmayan anlamına gelmez.

Öğrenme öğretim sürecinde bireylerde meydana gelen becerilerin uygulamada etkinliğini belirlemede ölçme ve değerlendirmeye ihtiyaç duyulmaktadır. Öğrencilerin psikomotor gelişimlerinin, bilişsel ve duyuşsal kazanımlarındaki deęişimlerinin görülmesi, başarısızlık gösteren öğrencilerin öğrenme eksikliklerinin giderilmesi, başarılı olanların güdülenmesi sürecinde ve öğretmenlerin kendilerini değerlendirmelerinde ölçme ve değerlendirme önemli yer tutar. Ayrıca ölçme ve değerlendirme sonuçları yöneticilere ve velilere önemli veriler sağlar (Karip ve dięerleri, 2009).

Değerlendirme, en genel anlamıyla ölçme sonuçları üzerinden yapılan yargılardır. Öğrenciye ait bir performans ya da ürün hakkında değer yargısında bulunulmasına değerlendirme denir (Başol, 2014). Öğrenci hakkında karar verme aşamasında bu yargılar önemli bir yer tutar. Ancak unutulmamalıdır ki bir hatadan kaynaklı yapılan bir değerlendirme bireyi olumsuzluęa sürükleyebilir. Özellikle değerlendirme yapılırken bireylerin davranışlarının gözlemlenmesi ve süreç içinde incelenmesi gerekir. Değerlendirmenin yapılabilmesi için öğrenmenin gerçekleşmiş olup olmadığı davranış formunda ölçülmüş olmalıdır (Tekindal, 2011).

Öğrencilerin davranışlarının gözlemlenip ölçülmesinde birey odaklı davranılmalıdır. Yani öğrenci merkeze alınıp inceleme yapılmalıdır. Öğrenci merkezli eğitim öğrencinin sadece not almasına odaklanmadığı gibi aynı zamanda dersin hedefleri doğrultusunda öğrenmesini ve etkin katılımını sağlar (MEB, 2007). Bu durum, eğitim sürecinde önemli olan ölçme ve değerlendirme uygulamalarına yüklenen anlamın deęişmesine sebep olmuş, bir yandan da sistem içinde kullanılan ölçme ve değerlendirme tekniklerinde de önemli deęişikliklerin oluşmasını beraberinde getirmiştir ( Bahar, Nartgün, Durmuş, ve Bıçak, 2008).

Ölçme ve değerlendirmenin önemi; eğitimin teorik bir bilim olma gayretleri yanında, pratik kararlarda değer yargılarına dayanak sağlamasından gelir (Turgut, 1984). Bu bakımdan ölçme ve değerlendirme eğitim sürecinin hem tamamlayıcısı hem de bir parçası olarak düşünölmelidir.

Ölçme ve değerlendirmenin yansız olarak yapılması; eğitim-öğretim çalışmalarının aksamasını, öğrencilerin başarısız ve mutsuz olmaları gibi olumsuz durumları ortadan kaldırır. Ayrıca öğrenci, öğretmen, idareci ve veliler arasında çözümlenmesi güç problemlerin ortaya çıkmasını önler ( Çağlar, 1970; Yıldırım, 1983; akt. Semerci, 1992). Ölçmenin yansız yapılması hata oranını düşürerek daha doğru sonuçlar elde edilmesini sağlar. Ayrıca bu durum bireyin gerek kendisiyle gerek çevresiyle çatışma durumuna düşmesine olanak vermez.

Eğitimde ölçme ve değerlendirme sürecinin önemli bir noktası da not verme işlemidir. Not verme, gözlenen öğrenciye ait ölçme sonuçlarının harflerle ya da sayılarla sınıfsal forma dönüştürme olarak ifade edilebilir. Not vermeyi amaçlayan kişi; bunu yaparken dikkatli olmalı ve bu işlemin gereklerini sağlayabilecek birikime sahip olmalıdır.

Eğitim-öğretimde notun gerekliliği, değerlendirmenin gerekliliğinden doğar. Değerlendirme daha çok bireylere verilen notlar üzerinden yapıldığı için not verme genel anlamı ile değerlendirme olarak düşünölmektedir. Cronbach; değerlendirmenin, bu sebeple de notların eğitimdeki fonksiyonunu şöyle ifade eder ( Semerci, 1992):

Değerlendirme, yeterince başarılı öğrenciyi güdüler.

Değerlendirme, öğrenci hakkında verilecek kararlara dayanak sağlar.

Değerlendirme, öğrenciye davranışını nasıl değiştireceğini ve nasıl geliştireceği hakkında bilgi verir.

Değerlendirme, öğretim elemanının kendi öğretimini ne derecede etkili olduğunu kestirmesine yardımcı olur.

Değerlendirme, yöneticilere ve diğer ilgililere bilgi verir.

Cronbach, yukarıda sıralanan maddelerde değerlendirmeye iki açıdan bakmıştır. Bunlardan birincisi ;sonuçlara bakarak çeşitli kararlar verme, ikincisi öğrenciye ve diğer ilgililere öğrenci başarısı hakkında bilgi vermedir. Thorndike ve Hagen da öğrenci

başarısı hakkındaki bilgilerin nerelerde kullanılabileceğini tartışmışlar ve yukarıdakine benzer bilgiler ortaya koymuşlardır (Turgut, 1984; Akt. Semerci, 1992). Yapılan araştırmaların ışığında değerlendirmenin dönüt boyutu da ortaya çıkmakta ve bu işlem bireyin gelişim sürecinin daha rahat ve daha ayrıntılı gözlemlenmesine olanak verip, çevrenin olumsuz etkilerinin de azaltılmasını sağlamaktadır.

Değerlendirme verilecek kararların ölçme süreçlerine göre ölçütün baz alınması zamanına göre kriter ve bağıl dayanaklı olarak yapılabilmektedir. Kriter dayanaklı bir yorum için, öncelikle testin de kriter temelli olması gerekir.

Ölçüt, ölçme aracı uygulanmadan önce tespit edilmiş olmalıdır. Kriter dayanaklı değerlendirmede, ölçme sonuçları önceden belirlenmiş kriterler ile karşılaştırılarak değerlendirme yapılır. Eğer ölçüt, ölçme aracı uygulandıktan sonra tespit edilmişse, değerlendirme norm dayanaklı olur. Norm dayanaklı değerlendirmede, bir öğrencinin performansı, diğer öğrencilerle karşılaştırılarak değerlendirme yapılır (Çelik, 2000).

## 2. 2. Ölçme Araçları

Eğitimde öğrenme çıktılarının miktarının belirlenmesi ve verilecek karara ulaşmada yol gösterecek olan notların bireylerden elde edilmesi ölçme araçları yardımı ile olmaktadır. Eğitim sürecinde öğrenme miktarının kestirilmesi ve sayısal olarak ifade edilmesinde pek çok ölçme aracı kullanılmaktadır. Ölçme araçlarını kullanacak kişilerin de bu araçları kullanırken özelliklerini çok iyi bilmeleri gerekmektedir. Çünkü kullanım amacı, mekanı, zamanı, bireylerin alt yapıları açısından kullanılacak ölçme araçları değişim gösterir. Ancak günümüzde ölçme aracı uygulayıcıları bireyin veya konunun uygunluğundan çok uygulama kolaylığına göre ölçme araçlarını seçmektedir. Özellikle bu süreçte birey sayısı, uygulama ve değerlendirme zamanı daha çok etkili olur.

Ülkemizde sürekli uygulanmakta olan ve bireylerin hayatlarını tamamen değiştirebilecek nitelikteki sınavlarda (YGS, SBS, LYS, ALES, YDS, TUS, KPSS,...) da bireylere ve seçim amacına uygunluk gibi niteliklerden çok zaman, maliyet ve uygulama kolaylığı daha önemli görülmektedir.

Daha öncede belirtildiği gibi eğitimde pek çok ölçme aracı bulunmaktadır. Sınıflandırılmasına bakıldığında ölçme araçları literatür incelendiğinde *klasik ölçme araçları* ve *alternatif ölçme araçları* olmak üzere iki gruba ayrılmakta ve bu şekilde incelenmektedir.

### **2. 2. 1. Klasik Ölçme Araçları:**

Geleneksel ölçme ve değerlendirme yöntemleri, öğretmenlerin çoğu tarafından bilinen ve eğitimin her kademesinde uygulanan yöntemlerdir. Bu yöntemler;

- 1- Yazılı yoklamalar
  - 2- Çoktan seçmeli testler
  - 3- Doğru yanlış testleri
  - 4- Eşleştirme soruları
  - 5- Kısa cevaplı testler
  - 6- Matematikte problem çözme testleri
- olarak gruplandırılabilir.

#### **2. 2. 1. 1. Yazılı Sınavlar**

Yazma becerisine sahip olunması yazılı sınavların en önemli özelliklerinden biridir. Bundan dolayı yazma becerisinin bu tür sorularda nasıl bir rol alacağı önemlidir. Bazı öğretmenler yazma becerisine önem verirken, bazılarının ise içeriğe önem verdiği görülmektedir. Yazma becerisi önemli olmasına rağmen asıl olan bilginin doğruluğu ve amaca uygunluğudur (Bahar ve ark., 2008).

Yazılı sınavlar; organize etme, ilişkilendirme ve iletişim becerilerinin ölçülmesinde çoktan seçmeli testlere göre büyük avantajlar sağlar (Aiken, 1991). Cevaplayıcının soruyla ilgili bilgiye sahip olmasının ötesinde problem çözme ve eleştirel düşünme gibi üst düzey bilişsel becerilerini kullanması gerekir (Haladyna, 1997). Ancak belli bir zaman içerisinde cevaplanması gereken soruların var olması yazılı sınavları bir güç testinden ziyade hız testine dönüştürür.

Yazılı sınavların özellikleri:



- 1- Cevaplayıcı vereceği cevabı kendisi düşünüp bulmalıdır.
- 2- Cevap yazılı olarak verilir.
- 3- Cevaplar tek düze değildir. Her cevaplayıcının kendi özgünlüğü vardır.
- 4- Cevapların tamamen doğru ya da tamamen yanlış olarak sınıflandırması mümkün değildir.
- 5- Cevapların doğruluk derecesinin tayin edilmesi puanlayıcıya bağlıdır.
- 6- Cevaplama işlemi cevaplayıcının çok zamanını alır.
- 7- Yazılı sınavların puanlaması oldukça güçtür.
- 8- Hazırlanması kolay ancak puanlaması oldukça uzun zaman alır.
- 9- Güçlük derecesini objektif yöntemlerle tayin etmek güçtür (Turgut ve Baykul, 2010).

Yazılı sınavın olumlu ve olumsuz yönleri aşağıda verildiği şekildedir.

#### Olumlu Yönleri:

- Hazırlanmasının kolay olması,
- Öğrencinin cevaplama özgürlüğüne sahip olması,
- Öğretmenin yorum sorabilme, öğrencinin yorum katabilme özgürlüğünün olması,
- Detaylı açıklama gerektiren konularda soru sorulabilmesi,
- Konunun gerçekten öğrenilip öğrenilmediği ve öğrenmenin derinliği hakkında bilgi vermesi,
- Üst düzey öğrenmelerin ölçülmesinde etkililik,
- Cevabı seçeneklerden seçmesi söz konusu olmadığından, şans başarısı faktörünün olmayışı,

#### Olumsuz Yönleri:

- Değerlendirmenin zaman alıcı olması,
- Değerlendirmenin sübjektifliğe açık olması,
- Puanlama güvenilirliğinin düşük olması,
- Soru sayısının sınırlılığından dolayı kapsam geçerliliğinin düşük olması olarak verilebilir (Başol, 2014).

### 2. 2. 1. 2. Çoktan Seçmeli Testler

Çoktan seçmeli sorular, bir soru kökü ve bunu izleyen bir seri muhtemel cevaptan ibarettir (Gronlund, 1985; akt. Semerci, 1992). Çoktan seçmeli sınavlar; değerlendirilmesi açısından objektif sınavlar olarak bilinmekte ve günümüzde eğitim alanında en sık kullanılan ölçme araçlarından. Özellikle ülkemizde de çok geniş katılımlı ve önemli sınavlarda (YGS, LYS, ALES, SBS, YDS vb.) değerlendirme açısından daha kolay olduğu için uygulama olarak tercih edilmektedir.

Soru kökü cevaplayıcıdan beklenen görevin ifade edildiği kısımlardan; seçenekler ise doğru cevap ve çeldiriciden oluşmaktadır. Doğru cevap, cevaplayıcıdan bilgisini kullanarak bulması düşünülen kısımlar; çeldiriciler ise doğru cevap dışında kalan ve doğru cevabın bulunmasını nispeten engelleyen, cevabı net olarak bilmeyenlere cazip gelebilen kısımlardır. İyi bir çeldirici cevap hakkında yeterince bilgisi olmayan cevaplayıcının cevaplama şansını azaltan özelliğe sahip olmalıdır. Diğer yandan yeterince bilgisi olan cevaplayıcının kolay eleyebileceği özellikte olmalıdır (Özçelik, 2011). Bu bakımdan çoktan seçmeli sınavların hazırlanması, gerek konu alanında gerekse dil ve anlatım alanında gerekli bilgiye sahip kişiler tarafından gerçekleştirilmelidir.

Çoktan seçmeli testlerin özellikleri aşağıda sıralanmıştır (Turgut ve Baykul, 2010):

1- Çoktan seçmeli testlerde öğrenciler cevabı yazılı vermediklerinden dolayı hem cevaplamanın süresi azalmaktadır hem de öğrencilerin yazı güzelliğinin ve yazı yazma kazanımlarının sürece katılması söz konusu değildir. Bu durum sınavın geçerliliğini yükseltir.

2- Öğrenciler cevabı yazmaya zaman ayırmadıkları için, sınav süresi içerisinde daha çok soru cevaplayabilmektedirler. Sınavda yoklanacak kazanımların sayısını yükselteceğinden, sınavın kapsam geçerliliği üst düzey kazanımların ölçülmesi hariç diğer sınavlara kıyasla çok yüksek olacaktır.

3- Öğrenciler, kazanımın ölçülmesinde, sorunun cevaplanmasını yazılı olarak vermediğinden dolayı öğrencilerin kazanımı gerçekten bilip bilmediği anlaşılamamaktadır. Çoktan seçmeli soruların hatırlama düzeyine yönelik becerileri

ölçtüğüne inanılır. Ancak iyi bir hazırlık ile çok yönlü kazanımın ölçülmesi mümkündür.

4- Soru sayısının fazla olması çoktan seçmeli testlerde okuma hızı gibi başka istenmeyen değişkenlerin sürece katılmasına sebep olur. Öğrencilerin okuma hızları birbirinden farklı olduğundan soru sayısının fazlalaşması okuma hızı yüksek olan öğrencileri avantajlı duruma getirebilir.

5- Bu tür sınavlarda soru hazırlanması diğer sınavlara kıyasla zordur ve zaman alır. Öğretmenlerin soru hazırlama becerisine sahip olması gerekir.

6- Çoktan seçmeli testlerin okunması oldukça kolaydır. Cevap anahtarı hazırlanması durumunda optik okuyucular yoluyla otomatik olarak puanlama yapılır.

7- Soruların cevaplarının seçeneklerde verilmesinden dolayı doğru işaretlendiğinde puanlama tam objektif yapılabilir. Kısmi puanlama imkanı bulunmadığından dolayı cevaplayana göre puanların değişimi söz konusu değildir. Kısaca çoktan seçmeli testlerde puanlama güvenilirliği tamdır.

8- Çoktan seçmeli testler üzerinden test ve madde istatistiklerini hesaplamak mümkündür. Matematiksel formülleri vardır. Bu formüller ile soruların ve sınavın güçlüğü, geçerliği, güvenilirliği ve diğer özellikleri sayısal olarak kestirilebilir. İstatistiksel olarak istenilen özellikte ölçme araçları hazırlanabilir.

9- Çoktan seçmeli testler farklı öğrenim düzeylerine uygulanabilir.

10- Çok değişik madde çeşitleri olduğundan kazanımların ölçülmesinde alternatif soru tipleri vardır.

11- Çoktan seçmeli testlerin cevaplamasında doğru cevabın bulunmasında şans faktörünün rolü bulunduğundan bu yönü ile çok eleştiri almaktadırlar.

Çoktan seçmeli sınavların hazırlanması özellikle uzmanlık gerektiren bir süreçtir. Bunun başlıca nedenleri arasında hazırlanan soru kökünün, sorunun altında yer alacak muhtemel cevaplardan herhangi birine cevaplayıcıyı açıkça yöneltmemesi, yanlış bilgelere yer vererek cevaplayıcının aklını karıştırmaması ve hazırlanan sınavın ayırıcı olmasının sağlanmak istenilmesidir. Çoktan seçmeli sınavlar, uzmanlar tarafından özellikle şans başarısı içermesi nedeniyle eleştirilmektedir. Şans başarısının etkisinin ortadan kaldırılması için bu tipteki sınavlarda genellikle düzeltme formülü yardımıyla çalışılmakta ya da sınav hazırlayıcıları tarafından şans başarısını düşürmek için farklı yöntemlerde kullanılmakta ve yenileri araştırılmaktadır.

Çoktan seçmeli testlerde, seçeneklerin doğru cevaba yakınlığına yani çeldiricinin güçlüğüne göre ağırlık puanları verilerek soruların güvenilirliğinin önemli derecede yükseldiğinin ancak geçerliğinin pek değişmediği görülmüştür (Turgut, 1984). Yapılan bu uygulama yakın geçmişten itibaren ülkemizde de ölçme ve değerlendirme uzmanları tarafından tartışılmış ve uygulanma sıklığı kısmen artış göstermiştir.

### **2. 2. 1. 3. Doğru Yanlış Testleri**

Doğru-yanlış testleri, cevaplayıcının verilen ifadeleri “doğru” veya “yanlış” olarak işaretlediği sınav türüdür. Dersin içeriğine ilişkin geniş bir kısmı örnekleyerek hazırlanabilmekte ancak ders materyallerinden hazır alınarak kullanılan ifadeler öğrencileri ezberle yöneltebilmektedir (Aiken, 1991). Doğru-yanlış testleri üst düzey bilişsel beceri gerektiren öğrenme ürünleri ölçülmesinde yetersiz kalmaktadır.

Özellikle bu tip testlerde kullanılan bazı maddelerin muğlak olması öğrencileri ikilem arasında bırakmakta ve yanlış yönlendirebilmektedir (Aiken, 1991). Özellikle sosyal bilimlerle ilgili disiplinlerde “doğru” veya “yanlış” şeklinde ayıramayacak bir bilgi yapısının varlığı bu tip testlerin kullanımını sınırlamaktadır.

Doğru yanlış testlerinin özellikleri aşağıda verilmiştir (Başol, 2012).

İlkokulun üç yılında tercih edilir.

Doğru-yanlış maddeleri çoktan seçmeli testler gibi objektif olarak puanlanabilir.

Şans başarısı en çok olan ölçme aracıdır.

Doğru-yanlış soru formatı kesinlikle doğru veya kesinlikle yanlış önermeler için kullanılmalıdır.

Anlatım becerisi, yazı güzelliği ve sayfa düzeni gibi durumlar puanlamaya yansımayaacağından objektiftir.

### **2. 2. 1. 4. Eşleştirme Soruları**

Eşleştirme maddeleri iki grup halinde verilen, birbiriyle ilgili veya birbirlerini tamamlayan ifadelerin eşleştirilmesine dayanan ölçme araçlarıdır. Özellikle semboller, tarihler, yazar eserleri gibi hatırlama düzeyindeki öğrenmelerin ölçülmesinde sıklıkla

kullanılan ve objektif olarak puanlanabilen ölçme araçlarındandır. Güvenirliği arttırmak için soruda seçenek sayısı verilen öncüllerden daha fazla tutularak öğrencilerin eleme yöntemi ya da şans başarısı ile cevap vermesi engellenmeye çalışılmaktadır. Üst düzey bilgilerin ölçülmesinde yetersiz kalması ise en büyük dezavantajdır.

#### **2. 2. 1. 5. Kısa Cevaplı Testler**

Kısa cevaplı testler; genellikle bir cümle içerisinde boşluklar veya soru kökü verilerek; bir konu hakkındaki sayı, sembol, bir kelime ya da birkaç kelimelik bilgilerin sorulduğu ölçme araçlarıdır. Genellikle alt düzeyde (özellikle hatırlama düzeyinde) ölçmelerde kullanılmakta olan objektif ölçme araçlarındandır. Hatırlama düzeyinde bilgileri ölçtüğünden üst düzeydeki kazanımların ölçülmesinde yetersiz kalır.

#### **2. 2. 1. 6. Matematikte Problem Çözme Testleri**

Öğrencilerin problem çözme becerilerini ölçmeye çalışan testlerdir. Uygulanması bakımından farklı formatlarda olabilmekte, uygulama ve üstündeki bilişsel basamaklarda kullanılmaktadır. Puanlayıcı görüşüne dayanıldığı için puanlama sürecinde objektifliğe dikkat edilmesi gereken bir sınav türüdür.

#### **2. 2. 2. Alternatif Ölçme Araçları:**

Klasik ölçme ve değerlendirme teknikleri olan, yazılı sözlü yoklamalar, kısa cevaplı testler, çoktan seçmeli testlerin bütün öğrenme süreçlerini ölçmede yetersiz kalmaktadır. Bu yetersizlikten dolayı eğitimde başka ölçme ve değerlendirme yöntemlerine ihtiyaç duyulmuştur (Turgut ve Baykul, 2010).

Alternatif ölçme ve değerlendirme, çoktan seçmeli testlerin de içinde olduğu geleneksel ölçme ve değerlendirme dışında kalan değerlendirmelerdir. Alternatif ölçme ve değerlendirme gerçek hayatla ilgilidir, öğrenciyi merkeze alır. Geleneksel ölçme ve değerlendirme yöntemlerinin aksine sadece ürün odaklı olmayıp sürecin değerlendirmesini de dikkate alır. Bu yönüyle öğrencinin yüksek düzeydeki düşünceleri, problem çözme ve yaratıcılığı ön plana çıkar. Birden çok cevap seçenekleri, süreç ve stratejiler öğretmen tarafından değerlendirilir. Değerlendirmedeki bu değişim

öğrencinin öğrendikleri ile ilgili sorumluluk almasına ve gurur duymasına sebep olur (Özçelik, 2011).

Alternatif ölçme ve değerlendirme teknikleri aşağıda belirtilmiştir:

1. Tanılayıcı dallanmış ağaç,
2. Yapılandırılmış grid,
3. Kelimle ilişkilendirme,
4. Portfolyo (öğrenci ürün dosyası),
5. Proje
6. Performans değerlendirme,
7. Gözlem,
8. Kavram haritası,
9. Görüşme.

Alternatif ölçme ve değerlendirme teknikleri ya hep ya hiç kuralına göre değerlendirme yapmaz. Öğrencinin başarısı ve başarısızlığını değerlendirirken performansını, ne bildiğini de değerlendirme süreci içerisine almakta ve bu amaçla, puanlama ölçeklerini ölçme sürecine katmaktadır. Puanlama ölçeği, öğrencinin bir kavrama, duruma ve olaya ilişkin bilgisini ortaya koyması veya ödevi yapmaya ilişkin olarak öğrencinin yeterlilik düzeyini belirlemeye yönelik bir puanlama sistemidir. Bu sistem öğrencinin zayıf ve kuvvetli olduğu alanlarla ilgili geri bildirim verme ve iyileştirme amaçlı öğretimi planlaması için de kullanılabilir (Bahar ve ark., 2008).

Alternatif ölçme araçları daha çok bireylerin performanslarını sergilemeye yöneliktirler. Bireylerin öğrendiklerini uygulamaları, yansıtmaları, ilişkilendirmeleri ve yeni ürünler üretmelerine elverişlidirler. Özellikle ülkemizde 2006 yılından itibaren uygulanmaya başlanan yeni öğretim programının temel felsefesi olan *yapılandırıcı* eğitim anlayışı ile bireylerin süreçsel gelişmelerinin sağlanmasında alternatif ölçme araçları önem kazanmıştır. Öğrencilerin bilgileri daha önce öğrenmiş oldukları ve yeni öğrendikleri bilgileri bütünleştirip kullanarak yansıtma ve üretme süreçleri klasik ölçme araçlarıyla yetersiz kalmış, bunun için alternatif ölçme araçlarının kullanımı eğitimde önemli bir hale gelmiştir.

Son yıllarda alternatif ölçme araçlarının kullanımında artış olsa da ülkemizde hala fiziki yetersizlik ve uygulamaya yatkınlıktaki sıkıntılar başta olmak üzere bazı nedenlerden dolayı bu tür ölçme araçları ya yanlış kullanılmakta ya da hiç kullanılmamaktadır. Ancak alternatif ölçme araçları bireylerin performanslarının belirlenip puanlanması ve değerlendirilmesinde oldukça yardımcı olmaktadır. Alternatif ölçme araçlarının çoğunda yapılacaklar bireylere yönerge veya açıklaması ile verilmekte, aynı zamanda bireylere klasik sınavların aksine geniş zamanlar verilmektedir. Bunun yanında performansın belirlenmesinde öğrencilere verilen yörüngelerde gözlenmek istenen davranışın hangi ölçütlere göre değerlendirileceği belirtilip açıklanır.

Performans ölçütünün belirlenmesi, ölçme sonuçlarının geçerliği ve güvenilirliği açısından gerekli ve hem öğrenci hem de öğretmen için önemlidir. Öğrenci bu sayede kendinden beklenenin ve hedefin ne olduğu ile ilgili uyarılmış olur. Ölçmede puanlama esasları belirlendiği için ölçme aracının geçerliği ve güvenilirliği yükselmiş olur (Turgut ve Baykul, 2010). Alternatif ölçme araçları ile değerlendirme yapılırken bu ölçütler başta verilmekte ve performansı gösterecek bireyden istenenler başta belirtildiği için ölçme sürecinin kalitesi yükselmektedir.

### **2. 2. 2. 1. Tanılayıcı Dallanmış Ağaç**

Tanılayıcı dallanmış ağaç; bir konu hakkında öğrencilerin bilgi seviyelerinin belirlenmesi, neyi bildiklerinin veya neyi bilmediklerinin ortaya çıkarılması için kullanılan bir tekniktir. Bu teknikte temelden ayrıntıya doğru bir sıralama yapılır. Öğrenciler kendilerine sunulmuş ifadelerle doğru ya da yanlış şeklinde cevap vererek seçimlerinin yönlendirdiği soruya geçerler. Buradaki amaç öğrencinin bu sorulara cevap vererek en doğru çıkışı bulmasıdır (Çepni ve Çil 2009).

Tanılayıcı dallanmış ağacı hazırlamak için 7 ila 16 arasında doğru veya yanlış şeklinde cevap verilebilecek önermeler yazılır. Tanılayıcı dallanmış ağaç birbirleriyle ilişkili konuları ölçmek için idealdir. Dallanma ne kadar çok adımda gerçekleşirse detay artacağından soru güçleşir. Önermeler en baştaki önerme en genel ve en uçtaki en özel olacak şekilde sıralanmalıdır (Başol, 2014).

Tanılayıcı dallanmış ağaç bireylerin yanlış yaptıkları noktada geri dönüp hatalı adımlarını düzeltmelerine imkan vermesi, bilgileri ilişkilendirmeyi sağlaması, doğru ve yanlış bilgileri ayırt etmesi açısından avantajlı fakat; şans başarısıyla cevabı bulma ihtimalinin yüksek olması, birbiriyle ilişkili, net, anlaşılır ve tutarlı önermeler yazılması açısından hazırlanmasının güç olması, üst düzey zihinsel becerilerin ölçülmesinde yetersiz kalması açısından ise dezavantajlı bir ölçme aracıdır.

### **2. 2. 2. 2. Yapılandırılmış Grid**

Yapılandırılmış grid tekniği ilk olarak Egan'ın (1972) çalışmalarından kaynaklanmış olup, daha sonra bazı araştırmacılar tarafından kullanılmış ve geliştirilmiştir (Johstone ve diğerleri, 2000; Akt. Özcan, 2011). Yapılandırılmış grid tekniğinde öncelikle bir tablo hazırlanır ve her bir gözeneğe sıra numarası veya harf verilerek gözeneklere tarih, isim veya tanımlar yerleştirilir. Tablonun altına gözeneklerden bir veya bir kaçını cevap olarak eşleştirilecek sorular yazılır. Birey tabloyu kullanarak bu soruların cevaplarını sorunun yanına yazar.

Bu teknikte; öğrencilerden bilgilerini mantıksal sıraya koyması veya seçmesi isteneceği için öğrencilerin konu alanına hakim olmaları gerekmektedir. Öğrenciler konunun sadece belli kısmını bilseler de her seçilen doğru kutucuk yani kısmi bilgi için puan alabilmektedirler. Uygulanma süreci ve değerlendirmesi kolay olması, öğrencilere kısa sürede uygulanabilmesi ve farklı yapısı ile dikkat çekmesi açısından avantajlı bir alternatif ölçme aracıdır.

Yapılandırılmış grid tekniğinin sınırlılığı ise tekniğin hazırlanması başlangıçta öğretmenler için biraz zahmetli olması ama zamanla pratik kazanılarak etkili bir biçimde kullanılabilmesidir ( Bahar ve ark., 2008).

### **2. 2. 2. 3. Kelime İlişkilendirme Testi**

Kelime ilişkilendirme testlerinde öğrencilerden, konuyla yakından ilgili temel kavramları seçmeleri ve seçtikleri bu kavramların zihinlerinde çağrıştırdığı kelimeleri sırasıyla yazmaları istenir. Her bir kavram için 20-30 saniye süre verilir ve verilen sürede öğrencilerin akıllarına gelen kavramları alt alta yazmaları istenir. Anahtar



kavramların birbirini etkilememeleri için her kavram ayrı sayfada ele alınır ve ardından öğrencilerin bu kavramları cümle içerisinde kullanmaları istenir. Anahtar kavramla öğrencinin ilişkilendirdiği kavram sayısına göre ilgili konunun ne derecede derinlemesine öğrenildiği hakkında fikir sahibi olunur (Başol, 2014).

#### 2. 2. 2. 4. *Portfolyo*

Portfolyo, bireylerin belirli bir alanda belli amaçlarla gerçekleştirdikleri çalışmaları, etkinlikleri kronolojik bir biçimde kapsayan ve o çalışma sürecinde bireyin performansının ve gelişiminin genel bir fotoğrafını ortaya koymayı amaçlayan kişisel dosya olarak düşünülebilir (Alıcı, 2011; ed. Tekindal, 2011). Portfolyo çalışmaları ülkemizde eğitim sistemiz içerisinde daha çok *ürün dosyası* olarak yer almaktadır. Portfolyo çalışmalarında, öğrencilerin süreç içinde yapmış olduğu bütün etkinliklere yer verildiği için öğrencinin gelişimi de gözlenebilmektedir. Dosyaya konulacak ürünler öğrenci tarafından seçilmekte ve öğrencinin yaptığı çalışmalardan hangilerinin yer alacağını kendisinin seçmesi, çalışmalarını kendisinin gerçekleştirmesi üretme ve başarı duygusunun da gelişmesine olanak sağlamaktadır.

Portfolyoların yararları verilecek olursa,

Öğrencinin süreçteki gelişimini göz önüne serer,  
Öğrencilerin kendilerine seçim ve katılım hakkı verir,  
Öğrencilerin üretme ve başarı duygularını tatmalarını sağlar,  
Üst düzey kazanımların elde edilmesine olanak sağlar,  
Öğrencilerin çok yönlü değerlendirilmesine olanak verir,  
Velilere ve öğretmenlere öğrenci hakkında bilgiler sunar,  
Öğrencinin anlık değil farklı zamandaki ve durumdaki düşüncelerini yansıtmasını sağlar.

Portfolyoların bunun yanında, değerlendirilmesinin zor, zaman alıcı ve uzmanlık gerektirmesi, saklanması ve puanlayıcı güvenilirliğinin sağlanmasının zor olması gibi olumsuz yanları da vardır. Portfolyo çalışmaları özellikle eğitim sistemimizdeki sınav ve program odaklı yapıdan dolayı gerekli önemi görememektedir.

### 2. 2. 2. 5. Proje

Öğrencilerin bireysel ya da grup halinde istedikleri bir alanda veya konuda inceleme, araştırma ve yorum yapma, görüş geliştirme, yeni bilgilere ulaşma, özgün düşünce üretme ve çıkarımda bulunmaları amacıyla ders öğretmeni rehberliğinde yapacağı çalışmaları ifade etmektedir (MEB, 2008).

Proje ödevleri, uzun sayılabilecek birkaç hafta ya da birkaç ayda tamamlanabilecek çalışmalardır. MEB yönetmeliğine göre proje ödevleri her öğrenci tarafından istediği dersten bir yıl boyunca en az bir adet alınmak zorundadır. Proje ödevleri öğrencilerin araştırma yapma ve bilgilerini kullanarak ortaya bir ürün koymasını amaçlayan süreçsel çalışmalardır.

Projeler, öğrencilerin araştırma ve inceleme, problem çözme, öğrendiklerini kullanma gibi üst düzey düşünme gerektiren, gerçek yaşama benzer işler üzerinde, özgün bir ürün ortaya koymak amacıyla yaptıkları çalışmalardır (Açıkgöz, 2004). Bununla birlikte öğrenciler grupça veya bireysel çalışarak, proje bitiminde yazılı, görsel veya sözel olarak sunumlarını yapmaktadırlar.

Ülkemizde proje ödevleri daha çok öğretmenlerin belirledikleri konularda ve yine öğretmenlerin seçtikleri öğrencilere verilmektedir. Uygulama açısından proje ödevlerini öğrenciler kendileri seçmeli ve öğretmenlerin rehberliğinde yapmalıdırlar. Bir başka hatalı uygulama da proje ödevlerinin, ödevlere yardımcı olmaları beklenen veliler tarafından bizzat yapılmasıdır. Ayrıca öğrencilerin buldukları gelişim dönemleri, mevcut durumları ve çevre şartlarına uygun olarak proje ödevleri seçilmelidir. Bunlara dikkat edilmemesi proje ödevlerinin güvenilirliklerini ve geçerliğini düşürmektedir.

Proje ödevlerinin bir başka olumsuz olarak algılanan yanının değerlendirmede yeterli özenin gösterilmemesi olduğu görülmektedir. Proje ödevleri değerlendirilirken daha çok analitik puanlama ölçeği tercih edilirse daha ayrıntılı ve objektif puanlama sağlanabilir.

### **2. 2. 2. 6. Performans Değerlendirme**

Klasik olarak yapılan ölçme ve değerlendirmenin temel amacı; öğrenme amaçlarının süreç sonunda ne kadar gerçekleştiğini ölçmektir. Günümüzde yapılan ölçme ve değerlendirme, öğrencilerin öğrenme sürecine ve ortamına katkı sunulmak üzere sahip oldukları yetenekleri yazılı, sözlü ve kazanım olarak göstermeyi amaçlamaktadır (Çepni ve Çil. 2009). Performans değerlendirme özellikle davranışsa, gözlemlerde ön plana çıkmaktadır.

Performans değerlendirmede bütün öğrenme ürünlerinin değerlendirilmesi söz konusudur. Bu ürünler; bilişsel, duyuşsal, ve psikomotor kazanımlarda üst düzey düşünme becerileri olarak adlandırılan akıl yürütme, problem çözme, iletişim becerileri; konferans verme, müzik aleti çalma, deney, spor yapma olarak tanımlanabilir (Turgut ve Baykul, 2010). Performansın gözlemlenip değerlendirmesi sürecinde, öğrencilerin çok boyutlu ve üst düzeyde düşünceleri ortaya çıkarılmaya çalışılmaktadır.

Ölçme ve değerlendirmenin öğrenci performansı üzerindeki etkisini inceleyen çalışmalarda ölçme ve değerlendirme etkinliklerinin öğrenmenin bir parçası olduğuna değinilmiştir. Brophy (1987), öğretmenlerin öğrencilere verdiği dönütün, öğrencinin sonraki öğrenme çabalarını yönlendirdiğini belirterek, öğrencinin konu başlıklarına ilişkin erişim düzeyi hakkında net bir bilgi edindiğini ve eksiklik gördüğü noktalara dikkatini verdiğini vurgulamıştır. Stipek (2002) ise öğrenci performansının değerlendirilmesiyle öğrencinin kendine duyduğu güvenin arttığını belirtmiştir. Buna ek olarak öğrenciler hangi yeteneklerinin geliştirilmesi gerektiğine dair fikir edindiklerinde, erişim düzeylerinin kontrol edildiği algısına sahip olmakta ve bunun da öğrencilerin öğrenmelerini güçlendiren bir etken olduğu ifade edilmiştir (Weiner, 1992; Stiggins, 1998).

### **2. 2. 2. 7. Gözlem**

Gözlem, planlı ve sistemli bir veri toplama yöntemidir. Öğrencilerin sınıf içi tartışmalara katılımı, sorulara verdiği yanıtlar, grup çalışmalarına katılım, fen bilgisi

dersinde yaptıkları deney çalışması, müzik aleti çalması gözlemlenebilen kazanımlardır ve bu tür kazanımları değerlendirirken gözlemden faydalanılır. (Özçelik, 2011).

Gözlem doğal ve yapay olmak üzere iki şekilde yapılabilir. Doğal gözlemden öğrenciye ve bulunduğu ortama müdahale edilmeden gözlem yapılır. Ancak yapay gözlemden, gözlem koşulları ve gözlem yapılan mekan kontrol altına alınabilir.

Gözlemler durum belirlemede kullanılan dayanakların başında geldiği için oldukça önemlidir. Gözlemler öğretmenin öğrenci hakkında izlenimlerin anekdotu şeklinde olabilir. Bu şekilde tutulan küçük notlar ,daha sonra birey hakkındaki büyük bir resmi tamamlamada yardımcı olur. Anekdotaların yanı sıra dereceli puanlama anahtarı ve kontrol listeleri gözlem yaparken kullanılır. Kontrol listesi ile eksikler saptanarak bunlara yönelik plan geliştirilip uygulamaya konulabilir ve bu sürece göre bireyin ilerlemeleri kayıt altına alınabilir (Başol, 2014).

İyi bir gözlem yapmak için ( Özçelik, 2011);

1. Gözlenecek kazanımlar belirlenmelidir.
2. Bu kazanımları ölçecek ölçekler hazırlanmalıdır.
3. Öğrencilerin belirlenen kazanımları göstereceği ortamlar oluşturulmalıdır.
4. Öğrenciler bu durumlarda gözlenmeli ve puanlanmalıdır.

#### **2. 2. 2. 8. Kavram Haritaları**

Ausubel' in öğrenme üzerine önceki bilgilerin etkisini vurgulayan anlamlı öğrenme teorisine dayalı, 1974 yılında Joseph Novak'ın Cornell Üniversitesi öğrencileriyle beraber yürüttükleri bir araştırma projesi sonucunda geliştirilen kavram haritaları, insanların bilgiyi nasıl öğrendiklerini ve nasıl anlamlandırdıklarını gösteren bir öğrenme-öğretme stratejisidir. “Bireyin sahip olduğu kavramlar ve önermeler ile yeni bilgileri ilişkilendirerek bilgiyi oluşturması” olarak tanımlanan anlamlı öğrenmede bilgiler yeniden organize edilir, yapılandırılır ve böylece zihinde yeni bir anlam oluşturulur, buna karşın ezber yönteminde bilgiler bellekte düzensiz olarak önceki bilgilerle ilişkilendirilmeden yerleştirilir (Ausubel, 1968; Novak, 1993; Kaya, 2003).

Brinkman (2003) kavram haritası hazırlanırken dikkat edilecek hususları aşağıdaki gibi sıralamıştır.

1. Büyük bir kağıt kullanın.
2. Başlığı haritanın en üstüne yerleştirin.
3. Daha genel ve soyut kavramlar en yukarıda, daha özel ve somut kavramlar daha aşağıda olacak şekilde kavramları değişik seviyelerde yerleştirin.
4. Eğer mümkünse kavramları bağlantı kurulacak şekilde yerleştirin.
5. Kavramların altındaki en son satıra ilgili birkaç örnek ekleyin.
6. İlgili kavramlar arasında üsttekenden aşağıdakine doğru çizgiler çizin. Aynısını aynı seviyedeki kavramlar arasında da yapın.
7. Bağlantı çizgilerine kavramlar arası ilişkiyi açıklamak amacıyla kelime veya kelime grupları yazın.
8. En son satırın altına orada belirtilen kavramla ilgili örnekler yazın ve bu örnekleri ait oldukları kavramlarla bağlayın. (Bağlama ifadeleri olarak “örneğidir” gibi kelimeler kullanılabilir).
9. Kavramları daire içine alın. Örnekleri daire içine almayın.

### **2. 2. 2. 9. Görüşme**

Özgüven'e (1980) göre görüşme, en az iki kişi arasında sözlü olarak sürdürülen bir iletişim sürecidir. Görüşme tekniği; bireylerin tutum, duygu, düşünce ve inançlarına ilişkin bilgilerin elde edilmesinde etkili bir yöntem olması nedeniyle sosyal bilgilerdeki araştırmalarda sıklıkla kullanılmaktadır. Görüşme tekniği; bilginin veya görüşün alınma durumuna göre farklı şekillerde yapılabilmektedir.

Görüşme tekniği yapılandırılmış, yarı yapılandırılmış veya yapılandırılmamış görüşme olarak üçe ayrılabilir. Yapılandırılmış görüşmelerde, görüşmenin yapıldığı her öğrenciye aynı sorular, aynı formatta sorulur. Bu tür görüşmelerde çoğunlukla açık uçlu sorulara yer verilmez. Yarı yapılandırılmış görüşmeler önceden planlanan sorulardan oluşmakla birlikte, görüşme sırasında cevapların ayrıntılı olarak verilmesine olanak sağlamak amacıyla, ek soruların da sorulması söz konusudur. Yapılandırılmamış görüşme ise, herhangi bir görüşme protokolü olmaksızın spontane gerçekleştirilen görüşme şeklidir (Tekindal ve ark., 2011).

Görüşme sürecinde dikkat edilmesi gereken bazı noktalar vardır. Görüşme tekniği sürecin kayıt altına alınmasına ve açıklamalar ile tepkisel durumların incelenmesine olanak vermesi , yüz yüze etkileşim sağlayıp rahatlatıcı ve daha açık bilgiler sağlaması gibi olumlu taraflarının yanında; görüşmeyi yapan kişinin cevapları kendi amaçlarına göre anlamasına neden olması , zaman, maliyet, süre ve soruların standart olmaması gibi olumsuz yönleri de bulunmaktadır. Bu açıdan görüşmeyi yapacak bireyin bu durumları bilerek ve dikkat ederek gerçekleştirmeye özen göstermesi gerekmektedir.

### 2. 3. Genellenebilirlik Kuramı

Genellenebilirlik (G) Kuramı ölçme sonuçlarının güvenilirliğinin belirlenmesini, güvenilir gözlemlerin tasarımı, araştırılmasını ve kavramsallaştırılmasını sağlayan istatistiksel bir kuramdır. Genellenebilirlik Kuramı, Klasik Test Kuramının bir uzantısıdır (Cronbach ve diğerleri, 1972; Brennan, 2001). G Kuramı, klasik güvenilirlik kuramının günümüzde hala popüler olan gerçek puan modelinin sınırlılıklarına olan tepkiler doğrultusunda Cronbach ve arkadaşları, (1963) tarafından ortaya atılmıştır. Klasik güvenilirlik kuramı, bir tek gerçek puana sahip her bir gözlem ya da test puanının paralel gözlemlerin bir grubuna ait tek bir güvenilirlik katsayısı üretmesi fikri etrafında merkezlenir (Lord ve Novick, 1968; Baykul, 2000). G Kuramı 50 yıldır ölçüm prosedürlerinin geliştirilmesi uygulanmış olmakla birlikte, özellikle eğitim araştırmaları içinde uygulanması sınırlı kalmıştır (Bottema-Beutel, Lloyd, Carter ve Asmus, 2014).

Klasik test Kuramında ölçüm hatası gerçek puanla gözlenen puan arasındaki farktır ve tek bir genel hata terimi adı altında ele alınır. Oysa ki hata terimi ( $e$ ) puanlayıcı hataları ve ölçme aracında kaynaklanan hatalar gibi farklı öğelerin birleşkesidir (Briesch, Swaminathan, Welsh, Chafouleas, 2014). Klasik Test Kuramı sistematik hatayı tam olarak bağdaştıramamakta, ancak tesadüfi hatayı açıklayabilmektedir. Bu açıdan da KTK daha çok bağıl değerlendirme yapmaya olanak sağlamaktadır. Ancak G Kuramı hem sistematik hem de tesadüfî hatayı açıklayabildiği için hem bağıl hem de mutlak değerlendirmeye olanak sağlamaktadır.

G Kuramı bir grup bireyden hatta bazen tek bir bireyden elde edilen ölçme sonuçlarının, bu sonuçların elde edildiği belli sayıdaki maddelerin, puanlayıcıların ya da

durumların çok daha ötesine genellenebilmesi amacını taşır (Brennan, 1992; Shavelson ve Webb, 1991). G Kuramı, KTK' nın genişletilmiş hali olarak kabul edilir ve bazı yönlerden KTK' nın sınırlıklarına çözümler üretir ; ölçmeye karışabilecek her hata kaynağını (birey, görev, puanlayıcı gb.) hem kendi başlarına, hem de birbiri ile kombinasyonu şeklinde ele alır ve kaynakların kombine etkisini hesaplar. Fakat KTK' da bu hata kaynakları tek olarak ele alınır, ortak etkileri olabileceği göz önünde bulundurulmaz (Brennan, 2001:4; Rentz, 1987).

KTK' da her bir değişkenlik kaynağı durumları için farklı hesaplamalar ile farklı katsayılar elde edilmektedir. Aynı anda birden farklı değişkenlik kaynağı için aynı analizleri içeren hesaplamalar ve yorumlamalar yapılamamaktadır. G kuramı ise aynı anda farklı hata kaynaklarını değerlendirmekte, bunların ayrı ve birleşik hata etkilerini hesaplamaktadır.

G kuramı araştırmacıya varyans analizi yoluyla farklı hata kaynaklarını bir arada kullanarak onların büyüklüklerini kestirmesini sağlar. G kuramında varyans bölümleri çeşitli test dizaynlarının genellenebilirlik katsayısını nasıl etkilediğini ortaya koymak üzere kullanılır ; bu sayede test dizyanında alınan kararların ölçüm hataları üzerindeki etkilerinin daha doğru bir tahminini yapmak mümkün olacaktır (Brown, 2005). Tek bir analizde bir ölçmedeki birden fazla hata kaynağının ele alınması, G kuramının avantajıdır (Shavelson ve Webb, 1991). Özellikle sosyal bilimlerde birey üzerinde yapılan araştırmalarda tek bir hata kaynağının incelenmeye alınması araştırmada eksikliklere neden olabilmektedir. Çünkü birey psikolojik yönü tam olarak çözülemeyen bir canlı olduğu için pek çok hata kaynağının etkisinde kalabilmektedir.

G kuramının temeli varyans analizi (ANOVA) üzerine kurulmuştur. Varyans analiziyle toplam varyans desendeki bağımsız değişkenlere bölünür. Böylece ölçme sonuçları farklı varyans kaynaklarına ayrılarak bireylerin ya da objelerin gözlenen puanlarının evren puanlarına (gerçek puanlarına) genellenebilmesi sağlanır (Kaya, 2011).

G kuramı kapsamlı bir çerçeve ve ölçme sonuçları için istatistiksel yollar ileri sürer. Aynı zamanda bu kuram test puanlarının ve puanlayıcılar arası tutarlılığın da bir ölçüsüdür (Brennan ve ark., 2003). KTK' ya dayanmasının yanında G kuramının en önemli katkısı farklı senaryolara da imkan vermesidir.

G kuramı her ne kadar KTK ve ANOVA kaynaklı ortaya çıkmış olsa da, onlardan daha yüksek bir anlayış gerektirir. G kuramı ne KTK'yı ne de ANOVA'yı tüm yönleri ile ele alır, öyle ki KTK'nın ve ANOVA'nın bazı yönleri ile tutarsızlık gösterir. Ancak ana hatları ile alacak olursak, G kuramı varyans bileşenleri ve tahminleri üzerine yoğunlaşmaktadır (Brennan, 2001:4).

Kriter referanslı testlerde sistematik ve tesadüfi ölçüm hataları göz önünde bulundurulmalıdır. Klasik test kuramı sistematik ölçüm hatasını bağdaştıramamaktadır bu nedenle sadece norm referanslı yapılan testler için uygundur. Halbuki; Genellenebilirlik kuramı, hem norm referanslı değerlendirme hem de kriter referanslı değerlendirme yapılan testlerden her ikisi için de uygundur çünkü bağıl ve mutlak ölçüm hatalarının ikisini de sağlamada esnekliğe sahiptir ( Suen ve Lei, 2007).

### **2. 3. 1. Facet**

Genellenebilirlik kuramında her bir değişkenlik kaynağı facet olarak adlandırılmaktadır ve bu değişkenlik kaynağının koşulları vardır. Örneğin, bir başarı testinde maddeler değişkenlik kaynağı olarak adlandırılırken testte bulunan her bir madde de bu değişkenlik kaynağının koşulu olarak düşünülebilir (Wang, 2005).

G çalışmasında testte verilen tasarlanmış ve yönlendirilmiş facetler için varyans bileşenlerini araştırırken dikkatli olunmalıdır. Yapılan çalışmalardaki bütün varyans kaynakları ele alınmalı ve varyans kaynaklar birbirleriyle karıştırılmamalıdır. Aksi halde farklı hata kaynakları, aynı kaynak olarak alınabilir. G Kuramı KTK' nın bir uzantısıdır, fakat G Kuramı'nın bunun yanında faydalarından biri de testte olabilecek tüm facetler için tahmini varyans bileşenlerini ayırıştırma yapmasına imkan vermesidir. (Kumazawa, 2009).

### **2. 3. 2. Çaprazlanmış ve Yuvalanmış Desen**

Genellenebilirlik kuramına göre değişkenlik kaynakları çapraz (crossed) ya da yuvalanmış (nested) şekilde olabilir (Rentz 1987). Burada çaprazlanmış desende değişkenlik kaynakları diğer değişkenlik kaynaklarının tüm koşullarıyla örtüşmektedir. Örneğin; bir sınavdaki bireylerin bütün maddeleri cevaplandırmaları ve bütün bireylerin



cevaplamış oldukları tüm maddeleri bütün puanlayıcıların puanlamaları halinde bu desen çaprazlanmış desen olmaktadır. Çaprazlanmış desende değişken kaynakları arasına 'x' işareti konulmaktadır. Yani verilen örneğe göre; birey (b), madde (m) ve puanlayıcı (p) değişkenlerimiz için araştırma desenimiz  $b \times m \times p$  şeklinde olacaktır.

Araştırmada bir değişken kaynağı diğer değişken kaynağının tüm koşulları ile örtüşmüyor sadece belli koşulları ile örtüşüyorsa bu çalışma desenine yuvalanmış desen denilmektedir. Örneğin; sınavda her bir maddeyi ayrı birey cevaplıyor ve her maddeyi ayrı puanlayıcı puanlıyorsa bu araştırma desenine yuvalanmış desen denilmektedir. Yuvalanmış desende değişkenlik kaynakları arasına ' : ' işareti konulur. Bu örneğimizi ele alırsak birey (b), madde (m) ve puanlayıcı (p) değişkenleri için araştırma desenimiz  $b : m : p$  olacaktır.

Tümüyle yuvalanmış desen, değişkenlik kaynaklarına ilişkin en az bilgi elde edilebilen desendir. Aynı zamanda bu desen ile yapılan G çalışmasına bağlı yapılabilecek tek bir K çalışması vardır ve bu K çalışmasında yer alan desen de tümüyle yuvalanmıştır.

### **2. 3. 3. Tesadüfi ve Sabit Değişkenlik Kaynakları**

Araştırmacı, araştırma sonuçlarını örneklem ötesine yani evrene genellemek istiyorsa burada değişken kaynağı tesadüfi değişken kaynağı olarak isimlendirilir. Ancak araştırmacı sonuçlarını evrene değil sadece örnekleme genellemek istiyorsa değişkenlik kaynağı sabit değişken kaynağı olarak isimlendirilir.

Örneğin; öğrencilerin dört farklı formdaki sınav türünden aldıkları puanları tüm sınav türlerinden aldıkları puanlara genellemek isteyen araştırmacı için araştırmanın değişkenlik kaynağı olan sınav puanları örneklemeden evrene genelleme yapılmak istendiği için tesadüfi değişkenlik kaynağıdır. Ancak dört formdaki sınav puanlarını bu dört forma genellemek isteyen araştırmacı için değişkenlik kaynağı olan sınav puanları sabit değişkenlik kaynağı olmaktadır. Bunun yanında şu nokta belirtilmelidir ki G çalışmasında en az bir yüz tesadüfi olmalıdır (Alharby, 2006).

### **2. 3. 4. G Çalışması ve K Çalışması**

G Kuramında, güvenilirliğin araştırılması iki aşamadan oluşmaktadır. Bunlardan ilki Genellenebilirlik çalışması (G-çalışması) ve ikincisi Karar çalışması (K-çalışması) dır (Kaya, 2011). G çalışması, ölçüm hatasını makul ve ekonomik olarak çok yönlü yalıtım ve tahmin etmek, uygulama yapabilmek için tasarlanmıştır (Shavelson ve Webb, 2005).

G çalışmasının amacı, ölçmenin birden çok kullanımını kestirmek ve bu sayede varyans kaynakları ile ilgili mümkün olan en çok bilgiyi sağlamaktır. G çalışması, mümkün olan en çok değişkenlik kaynağını içerecek biçimde tasarlanmalıdır. Yani G çalışması, kabul edilebilir gözlemlerin evrenini mümkün olan en geniş şekilde tanımlar (Shavelson ve Webb, 1991).

G-çalışması sürecinde, örneklemin evrene genellenebilmesi için, puanların değişkenliğinin tüm kaynakları (varyans bileşenleri) ve bunlar arasındaki etkileşimler aynı anda ANOVA yöntemi kullanılarak kestirilmektedir. Kestirilen bu varyans bileşenleri bir sonraki aşama olan K-çalışmasında kullanılmaktadır ( Kaya, 2011). G çalışması sonucunda elde edilen sonuçların K çalışmasında kullanımı söz konusudur ya da araştırmacı isterse buradan devam etmeyip çalışmasını G çalışması ile sonlandırabilir.

K-çalışması, karar vermek üzere belirli bir amaç için veri toplanan çalışmadır ve yapılan bir K çalışmasında, incelenen bireyleri tanımlamak için veri toplanabilir (Kaya, 2011). Bir G çalışmasına karşılık birden fazla K çalışması yapılabilir. K çalışması ile güvenilirlik katsayısına benzeyen genellenebilirlik katsayısına (*G katsayısı*) ve güvenilirlik indeksine (*Phi katsayısı* ) ulaşılır. G katsayısı evren puanı varyansının kendisi ile bağlı puan varyansının toplamına oranıdır ve bağlı modellerde çalışılmaktadır (Çakıcı Eser, 2011).

G katsayısı KTK' daki güvenilirlik katsayısına benzemektedir. G katsayısı, göreceli karar modelinde gerçek varyansın, göreceli varyans ve gerçek varyansın toplamına bölünmesi ile bulunmaktadır. Öte yandan güvenilirlik endeksi ya da Phi ( $\Phi$ ) katsayısı mutlak karar modeli ile kullanılmaktadır. Phi katsayısı, gerçek varyansın, mutlak hata varyansı ve gerçek varyansın toplamına bölünmesi ile hesaplanmaktadır. Diğer bir

deyişle, bu iki katsayı hatanın ne kořullarda kabul edileceğine göre farklılık göstermektedir (Alharby, 2006). Sonuç olarak, tek bir G-çalışmasından elde edilen aynı varyans kestirimlerine dayalı pek çok K-çalışması düzenlenebilir. K-çalışmasında kullanılan formül Spearman-Brown formülüne benzerdir (Mushquash ve O'Connor, 2006).

Evren puanı amaçlanan deęişkenlik kaynakları için kabul edilebilir gözlemlerin evreninden elde edilecek puanların ortalaması olarak bulunan, ölçme puanı olarak tanımlanır. Evren puanları varyansı, klasik test kuramında yer alan gerçek puan varyansına benzer. Klasik test kuramından farklı olarak, G kuramında iki ayrı hata varyansı bulunur. Bu farklılık, G kuramında iki ayrı anlamda karar vermenin mümkün olmasından kaynaklanmaktadır ( Kaya, 2011).

G kuramında baęıl ve mutlak deęerlendirmeler arasında bir ayırım söz konusudur. Bu nedenle güvenilirlik hesaplanırken hata varyansları arasında baęıl ve mutlak hatalara göre farklılık vardır. Fakat klasik test kuramında baęıl ve mutlak hatalar arasındaki ayırım göz önüne alınmadan güvenilirlik hesaplanmaktadır (Brennan, 2001).

Baęıl deęerlendirme, ölçme işleminin sonucunda elde edilen puanların, önceden belirlenmiş deęişmez bir ölçüt dışında başka ölçütlerle de kıyaslanıp deęerlendirme yapılmasıdır. Mutlak deęerlendirme ise, ölçümlerin kıyaslanacağı ölçüt uygulamadan önce belirlenmiş mutlak bir ölçüte dayanılarak yapılan deęerlendirmedir (Gültekin, 2005).

Genellenebilirlik kuramı çerçevesinde, hangi modelin uygun olduğu daima uygulama amacına yani arařtırmacının ölçme sonuçlarını nasıl kullanacağına baęlıdır. Eđer bir öğrencinin bir sınavdan geçmesi veya kalması sınavı alan diđer öğrencilerin o sınavdaki performanslarına baęlı ise arařtırmacı göreceli model kullanmalı; eđer öğrencinin sınavdan geçmesi veya kalması diđer öğrencilerin o sınavdaki performanslarından baęımsız olarak sadece kendi performansına baęlı ise arařtırmacı mutlak model kullanmalıdır (Kaya, 2011).

Bireylerin başarılarının belirlenmesinde konunun, gözlemin yapılacağı ortamın, bireylerin özelliklerinin gibi farklı durumlar gözetilerek farklı ölçme araçları

kullanılmaktadır. Eğitimde ölçme araçları kullanılırken genellikle bireyin veya konunun özelliklerinden ziyade kullanışlılık, zaman, maliyet ve uygulama kolaylığı daha çok ön plana çıkmaktadır. Bu nedenlerden dolayı da klasik ölçme araçları, ölçme süreç ve kuramları daha çok ön plana çıkmaktadır. Oysa, alışlagelmiş yöntemlere göre G ve D çalışmaları araştırmacılara; ölçme araçlarından elde ettikleri verileri değerlendirerek kendi amaçları ve ilgileri bağlamlarında ek veri toplamalarına gerek kalmadan senaryolar yardımı ile maliyet gibi açılardan fayda sağlamaktadır (Bottema-Beutel ve diğerleri, 2015 ).

Öğrenci başarısının belirlenmesinde daha çok öğrencilerin verilen süredeki bilgilerini ve becerilerini kullanarak performanslarını en üst düzeyde ortaya çıkarmaları beklenmektedir. Öğrencilerin göstermiş oldukları performanslar uzmanlar tarafından gözlenip değerlendirilmekte ve böylece performans başarı olarak ifade edilecek notlara çevrilmektedir. Tabi ki bu süreçte performans belirlemede, farklı ölçme araçlarının yardımına başvurulmaktadır.

Türk Eğitim Sisteminde, performans değerlendirme birçok durumda yazılı sınavlarla yapılmaktadır. Merkezi sınavlar ise; Bloom ve Piaget'in bilişsel kuramına göre yapılmaktadır. Bu sınavlarda öğrencinin belli sürede belli sayıdaki sorulara verdiği cevap sayısına bakılarak ölçme yapılır. Ancak bu tür sınavlarda öğrencinin üst düzey kazanımlarını ve öğrencinin grup içindeki performansını ölçmek oldukça zordur.

Objektif testlerde gerçek bilgilere ulaşmada iyi yapılandırılmış testler kullanılırken, performans değerlendirme de ise çok daha az yapılandırılmış ve problemi tanımlama, veri toplama ve organize etme, birleştirme ve değerlendirme yönü ön planda tutulur.

Performans değerlendirme, bilimsel yöntemin ve bilim adamlarının bir problemi çözmeye çalışırken takip ettikleri süreçlerin önem kazandığı, farklı materyal ve yaklaşımların denendiği problem çözme becerilerini içermektedir. Performans değerlendirmede tek doğru veya en iyi cevap diye bir şey yoktur. Bu özelliği ile diğer ölçme araçlarından ayrılır.

Problem çözmeye farklı performansların ve problem çözmeye yaklaşımlarının kullanılması eğitimde istenen bir durumdur. Bilim adamlarının herhangi bir problemi çözerken takip ettiği yolları içeren bir konu öğrenimi, sadece ürün ve cevaba odaklı bir öğrenimden daha etkilidir (Çepni ve ark. 2009, Akt. Özçelik, 2011).

## **2. 4. İlgili Araştırmalar**

Bu bölümde araştırmanın kapsamını oluşturan farklı ölçme araçları ve genellenebilirlik kuramı ile ilgili yurtiçi ve yurtdışında yapılan çalışmalara yer verilmiştir.

### **2. 4. 1. Genellenebilirlik Kuramı İle İlgili Araştırmalar**

Atılğan (2004), “Genellenebilirlik Kuramı ve Çok Değişken Kaynaklı Rasch Modelinin Karşılaştırılması” adlı çalışmasında 2002 ve 2003 yıllarında yapılan Müzik Öğretmenliği Özel Yetenek Sınavı verileri üzerinde Genellenebilirlik Kuramı ve Çok Değişken Kaynaklı Rasch Modeli analizleri uygulamış ve sonuçları karşılaştırmıştır. Genellenebilirlik Kuramı çerçevesinde görev ve puanlayıcı değişkenleri incelenmiştir. Elde edilen analiz sonuçlarına göre Genellenebilirlik Kuramı ve Çok Değişken Kaynaklı Rasch Modeli yaklaşımıyla elde edilen sonuçların kısmen tutarlı sonuçlar verdiği belirtilmiştir.

Yelboğa (2007), ”Klasik Test Kuramı ve Genellenebilirlik Kuramına Göre Güvenirliğin Bir İş Performansı Ölçeği Üzerinde İncelenmesi” adlı çalışmasında 2005-2006 yılları arasında ‘İş Performansı Ölçeği’ kullanarak Klasik Test Kuramı ve G Kuramından elde edilen güvenilirlik katsayılarını karşılaştırmıştır. Elde edilen sonuçlara göre, ölçek üzerinden elde edilen katsayıların G Kuramı ve KTK’ ya göre uyumlu sonuçlar verdiği bulunmuştur.

Güler (2008), Klasik Test Kuramı, Genellenebilirlik Kuramı ve Rasch Modeli üzerine yaptığı çalışmada matematik başarısının ölçülmesinden elde edilen puanların güvenilirliklerini hesaplamış ve üç kuramdan elde edilen sonuçları karşılaştırmıştır. Araştırma bulgularına göre, matematik başarısını ölçmek için kullanılan ölçme aracının güvenilir sonuçlar verdiği ve puanlayıcıların birbirleriyle uyumlu puanlamalar yaptığı belirlenmiştir. Ayrıca araştırmada, puanların güvenilirliğinin belirlenmesinde elde edilen

puanların kullanım amacına bağılı olarak üç kuramdan hangisinin tercih edilmesi gerektiğinin belirlenmesi ve en az iki kuramdan yararlanılması gerektiği belirtilmiştir.

Nalbantođlu (2009), ‘Performans Ölçümlerinde Genellenebilirlik Kuramıyla Farklı Desenlerin Karşılaştırılması’ adlı çalışmasında 2007-2008 eğitim yılında 48 öğrencinin, verilen görevlere göre puanlayıcıların birlikte ve dönüşümlü olarak puanlamaları üzerinde karşılaştırılması yapılmıştır. Yapılan analizler sonucunda puanlayıcıların birlikte ve dönüşümlü olarak yaptıkları değerlendirmelerde değişkenler için kestirilen varyans değerlerinin birbirleriyle paralellik gösterdiği görülmüştür. Ayrıca araştırmada puanlayıcılar arasında puanlamalarda farklılık olmadığı sonucuna ulaşılmıştır.

Çakıcı Eser (2011), ‘Genellenebilirlik Kuramı ve Lojistik Regresyona Dayalı Hesaplanan Puanlayıcılar Arası Tutarlığın Karşılaştırılması’ adlı çalışmasında öğrencilerin PISA sınavı sorularından oluşan bir ölçme aracı üzerinden performanslarının puanlanmasında, ortaya çıkan puanlayıcı tutarlığı belirlenip karşılaştırılmıştır. Analiz sonuçlarına göre lojistik regresyon analizi ve genellenebilirlik kuramının, puanlayıcılar arası tutarlığı belirlemede paralel sonuçlar ürettiği, ancak lojistik regresyon analizinin Genellenebilirlik Kuramı kadar hassas çıktılar vermediği ve Genellenebilirlik Kuramı’ na göre daha yüzeysel bir istatistik olduğu sonucuna ulaşılmış.

Kaya (2011), ‘Genellenebilirlik Kuramının Doldurma Kavram Haritası Değerlendirme Çalışmasına Uygulanması’ adlı çalışmasında iki farklı doldurma kavram haritası tekniğinin uygulama sıralarını değiştirerek ikili uygulamalar yapmıştır. Uygulama sonuçları üzerinde G Kuramı uygulanmış ve varyans bileşenlerinin değişimlerine bakmıştır.

Taşdelen (2009), ‘Nedelsky ve Angoff Standart Belirleme Yöntemlerinin Genellenebilirlik Kuramı ile Karşılaştırılmasına İlişkin Bir Araştırma’ adlı çalışmasında iki farklı standart belirleme yöntemi ile elde edilen kesme puanları, Genellenebilirlik Kuramı ile karşılaştırmıştır. Çalışmada SBS ‘de yer alan 16 Fen ve Teknoloji sorusunun 40 uzman tarafından puanlanması ile analiz için gerekli veri seti oluşturulmuştur. Yapılan Genellenebilirlik çalışmasında her iki yöntemde de, maddeler arası

farklılıkların ortaya çıkarılabildiği ve puanlayıcılar arasında bir tutarlılığın olduğu görülmüştür. Standart belirleme yöntemlerinin tek olarak ele alındığı Genellebilirlik çalışmasında ise; Nedelsky yöntemi ile yapılan puanlamada maddeler arası farklılığın ortaya çıkarılabildiği, ancak puanlayıcılar arasında kesme puanı belirlemede uyum olmadığı görülmüş; Angoff yöntemi ile yapılan puanlamada maddeler arası farklılığın ortaya çıkarılabildiği ve puanlayıcılar arasında bir tutarlılığın olduğu görülmüştür.

Güler (2009),’in ‘Genellebilirlik Kuramı ve SPSS ile GENOVA Programlarıyla Hesaplanan G ve K Çalışmalarına İlişkin Sonuçların Karşılaştırılması’ adlı çalışmasında Genellebilirlik Kuramı uygulamalarında kullanılmakta olan iki farklı paket program için uygulama sonuçları karşılaştırılmıştır. Ayrıca çalışmada G Kuramı terminolojisinden bahsedilmiş ve uygulama basamakları açıklanmıştır. Elde edilen sonuçlara göre iki farklı paket programla bulunan sonuçlar neredeyse eşit çıkmıştır.

Anıl ve Büyükkıdık (2012), ‘Genellebilirlik Kuramında Dört Facetli Karışık Desen Kullanımı İçin Örnek Bir Uygulama’ adlı çalışmalarında G Kuramı uygulayarak çok sayıda değişkenlik kaynağını ele almışlardır. Veriler üzerinde Genellebilirlik ve Karar çalışması yapılmış ve sonuçlara göre en büyük etkiye sahip değişkenin birey olduğu ve bunu puanlayıcı değişkeninin takip ettiği belirtilmiştir.

Nalbantoğlu Yılmaz ve Uzun Başusta (2012), ‘Genellebilirlik Kuramıyla Dikiş Atma ve Alma Becerileri İstasyonu Güvenirliğinin Değerlendirilmesi’ adlı çalışmalarında tıp fakültesi öğrencilerinin dikiş atma ve alma becerilerine ait performans puanlarının güvenilirliğini belirlemek ve puanlama güvenilirliğini öğrenci, puanlayıcı, beceri ve görev etkileşimlerini de dikkate alarak dengelenmemiş verilerde genellebilirlik kuramıyla incelemiştir. Uygulamada puanlayıcıların her biri dönüşümlü olarak ve birbirlerinden farklı sayıda öğrenciyi puanlamışlardır. Elde edilen sonuçlara göre, puanlayıcı değişkenliğinin puanlama farklılığına neden olmadığı bulunmuştur. Öğrencilerin dikiş atma ve alma becerilerine ait performanslarının değerlendirilmesi sürecine yönelik Genellebilirlik Kuramı ile elde edilen G ve Phi katsayıları ise kabul edilebilir düzeydedir..

Delicelioğlu (2012)' nun 'Futbol Yetilerine İlişkin Dereceleme Ölçeğinin Güvenirliğinin Genellenebilirlik Kuramına ve Klasik Test Kuramına Dayalı Olarak Karşılaştırılması' adlı çalışmasında futbolculara uygulanan ve futbolcuların teknik yetilerinin tespit edilmesinde kullanılan dereceleme ölçeğinden elde edilen veriler üzerinden KTK ve G Kuramına göre güvenilirlik katsayıları hesaplanmış ve karşılaştırılmıştır. G Kuramında güvenilirlik katsayısı olarak G ve Phi katsayılarının hesaplanmasında yer alan bağıl ve mutlak hata varyanslarını azaltabilmek için alternatif K çalışmaları düzenlenmiştir. Elde edilen sonuçlara göre, KTK kuramında hesaplanan değişkenlere göre farklı güvenilirlik katsayılarının elde edilmesi bir sınırlılık olarak görülmüş ve G kuramının bu açıdan tek bir güvenilirlik sayısı vermesi ile bu sınırlılığı ortadan kaldırdığı vurgulanmıştır.

Alkan (2013), PISA 2009 Okuma Becerileri performansını incelediği çalışmasında; değerlendirmede kullanılan açık uçlu soruların, birden fazla puanlayıcı tarafından birlikte ve dönüşümlü olarak puanlanmasıyla elde edilen farklı desenleri G Kuramına göre karşılaştırmıştır. Araştırmada, iki farklı senaryo üzerinden genellenebilirlik kuramında kullanılmak üzere iki desen tasarlanmıştır. Analizler EduG 6 Programıyla gerçekleştirilmiştir.  $\bar{o} \times s \times p$  ve  $(\bar{o}:p) \times s$  desenleri karşılaştırıldığında,  $(\bar{o}:p) \times s$  deseni ile kestirilen bağıl ve mutlak hata varyanslarının  $\bar{o} \times s \times p$  desenine göre daha küçük olduğu, dolayısıyla G ve Phi katsayılarının daha büyük değerler aldığı görülmüştür. İki desende yapılan Karar çalışmaları incelendiğinde, her iki desende de puanlayıcı sayısını artırmanın G ve Phi katsayılarında artış sağladığı görülmüştür. Araştırma sonuçlarına göre bazı kitapçıklarda puanlayıcı sayısını yarıya indirerek ya da madde sayısını yarıya indirerek, çoğu zaman her ikisinin sayısını da yarıya indirerek kabul edilebilir düzeylerde G katsayısına ulaşmanın mümkün olduğu sonucuna ulaşılmıştır. Bazı kitapçıklarda ise madde sayısı sabit tutularak, puanlayıcı sayısının artırılmasının zaman, işgücü ve ekonomi açısından uygun olacağı sonucuna ulaşılmıştır.

Yıldıztekin (2014), yapmış olduğu çalışmasında; matematik dersinde öğrencilere uygulanan ve problem çözme becerisini ölçen açık uçlu sorular, analitik ve bütünsel dereceli iki ayrı puanlama anahtarı kullanılarak 5 farklı matematik öğretmeni tarafından puanlanmıştır. Elde edilen puanların KTK ve G Kuramına göre güvenilirlik kestirimleri yapılmış ve puanlayıcılar arası tutarlık dereceleri belirlenmeye çalışılmıştır.



Araştırmanın verileri elde edildikten sonra Klasik Test Kuramı'ndan, Pearson momentler çarpımı korelasyon katsayısı, Spearman sıra farkları korelasyon katsayısı, Cronbach Alpha, Kappa ve Krippendorff Alpha katsayıları ile Genellenebilirlik Kuramı'ndan b x m x p çapraz deseninde değişkenlik kaynakları ve yüzdeleri belirlenerek güvenilirlik analizleri yapılmıştır. Araştırma sonucuna göre, KTK ve G Kuramı 'na göre elde edilen güvenilirlik katsayıları birbirine paralel ve oldukça yüksek bulunmuştur. Puanlayıcılar arası tutarlık düzeyinin yüksek olduğu ve analitik dereceli puanlama anahtarı ile elde edilen puanların tutarlığının bütünsel dereceli puanlama anahtarı ile elde edilen puanların tutarlığından göreceli olarak daha yüksek olduğu sonucuna ulaşılmıştır.

Clark 2008' de yapmış olduğu çalışmada, 102 öğrencinin matematik dersindeki durumlarını 15 saniyelik aralıklarla bireylerin ortaya koydukları ürünleri gözlemleyip, gözlem sonuçlarını kaydetmiş ve bu veriler üzerinde genellenebilirlik kuramı ile analizler yapmıştır. Araştırmada öğrencilerin aralıklarda gözlenme ve tek gözlenme durumları karşılaştırılmıştır.

Moon (1995)' un yaptığı çalışmanın amacı, bir performans değerlendirmeden elde edilen sınırlı sayıda gözlemin, gözlenemeyen olası durumlara uygun şekilde genellenip genellenemeyeceğini araştırmaktır. Öğrencilerin performanslarına ilişkin nasıl güvenilir ölçmeler yapılabileceği ve kağıt-kalem testlerine karşı performans değerlendirmelerden nasıl farklı bilgiler elde edilebileceği açıklanmıştır. Yapılan bu çalışmada, üniversite öğrencilerine istatistik dersine ilişkin hem performans değerlendirme, hem de kağıt-kalem testi birlikte uygulanmıştır. Performans değerlendirmenin güvenilirliğinin kestirimi ve desenin geliştirilmesi için genellenebilirlik çalışması yapılmıştır. Puanların güvenilirliği Klasik Test Kuramı ve Genellenebilirlik Kuramı ile analiz edilmiştir. Her iki ölçme aracına ilişkin göreceli bilgiler elde etmek için puanların arasındaki ilişkiyi belirlemek için korelasyon hesaplaması ve açıklayıcı (exploratory) faktör analizi yapılmıştır. Çalışmanın sonucunda, performansları puanlayan iki puanlayıcının önemli bir hata kaynağı olmadığı bulunmuştur. Ayrıca, öğrencilerin performanslarındaki tutarsızlığın başlıca kaynağının görevler olduğu gözlenmiş ve görev sayısının arttırılmasıyla puanların güvenilirliğinin arttırılabileceği sonucuna ulaşılmıştır. İki puanlayıcı tarafından elde edilen tüm puanlar arasındaki korelasyon ve G çalışmasının sonuçları, puanlayıcıların tutarlı bir şekilde öğrenci

performanslarını değerlendirebildiklerini göstermiştir. Araştırmanın sonucunda puanlayıcı sayısının bir azaltılması ve böylece genellenebilirlik çalışmasındaki desende yer alan puanlayıcı değişkenlik kaynağının çıkarılması önerilmektedir. İki ölçüm arasında oldukça yüksek korelasyon bulunmuş; faktör analizi sonuçları faktör yapıları ile madde ayırıcılıkları arasında bir ilişki olduğunu göstermiştir.

Burns (1998), yapmış olduğu çalışmasında G Kuramının farklı hata kaynaklarını ele alması açısından üstünlüklerinden bahsetmiş ve hata oranının azaltılması açısından belirlenen farklı koşullara göre çareler sunduğunu belirtmiştir. GK' nın bir yorumu olarak bu çalışmada gözlenen bir durumun pratik bir uygulaması tartışılmıştır. Araştırma sonuçlarına göre; gözlemlerinde ya da anketlerinde kullandıkları maddeleri ve ya planlama çalışmalarını geliştirmek için G Kuramı tarafından sunulan avantajları kullanarak değerlendirmenin daha avantajlı olduğu sonucuna varılmıştır.

Neill ve Neill (2015), yaptıkları araştırmada yedi farklı kantitatif duyuşal ağrı testini beş farklı değerlendirici yardımı ile farklı sezonlarda ve farklı zaman aralıklarında yaptıkları gözlemler üzerinde sonuçları KTK ve G Kuramına göre karşılaştırmışlardır. Elde edilen sonuçlara göre tek bir test yerine farklı testlerden elde edilen verilerin güvenilirliği ve fizibilite değerlerini yükselteceği gözlemlenmiştir. Ayrıca bu çalışmada; farklı değerlendiricilerin vermiş oldukları gözlem değerleri güvenilirlik katsayısının artmasında etkili olmuştur; sonuç olarak, birden fazla değerlendirici olmasının olumlu gelişme sağladığını ortaya konmuştur.

Lane ve Sabers (1989), öğrencilerin yazma becerilerini değerlendirmede puanlama sisteminin güvenilirliğini belirlemek amacıyla genellenebilirlik kuramını kullanmışlardır. Araştırmada 15 öğrenciye bir konu verip bu konu hakkında kompozisyon hazırlamaları istenmiştir. Öğrencilerin kağıtları sekiz öğretmen tarafından, bir puanlama sistemi oluşturularak değerlendirilmiştir. Öğretmenler puanlama sistemini dört kategoride hazırlamıştır. Araştırmada öğrenciler (s), puanlayıcılar (r) ve puanlama sistemindeki kategoriler (c) ile gösterilmek üzere kategori değişkeni sabit olarak belirlenip tüm değişkenlerin çapraz tasarlandığı s x r x c deseniyle G ve K çalışmaları yapılmıştır. Araştırma sonuçlarına göre puanlayıcı sayısının azaldıkça mutlak ve bağıl hata varyanslarının arttığı, G ve Phi katsayılarının ise azaldığı tespit edilmiştir.

Alkahtani (2012), yaptığı çalışmada öğrencilerin Kur'an öğrenmedeki performanslarını iki rubrik ile G Kuramı ve ÇDRM ile değerlendirmiştir. Araştırmada, 93 öğrencinin performansı üç farklı puanlayıcı tarafından iki farklı rubrik kullanılarak değerlendirilmiştir. Yapılan G Kuramı analizi sonuçları analitik rubrik ile yapılan değerlendirmeye ait katsayıların daha yüksek olduğunu ve öğrencilerin değerlendirilmesinde daha güvenli sonuçlar verdiğini ortaya koymuştur. Bulunan sonuçlara göre öğrencilerin Kur'an öğrenmedeki performanslarının değerlendirilmesinde analitik puanlama anahtarı kullanımının yaygınlaştırılması gerektiğinin üzerinde durulmuştur.

Au, Prahardi ve Shiell (2008), çalışmalarında ekonomik değerlendirmeler için hazırlanmış iki ölçeğin güvenilirliğini araştırmışlardır. Ölçeklerden ilki 16 maddelik Likert tipi bir ölçek, ikincisi ise 57 maddelik bir kontrol listesidir. Çalışmada 1990 ve 2003 yılları arasında yapılan 30 makale kullanılmıştır. Bu makaleler her iki ölçekle iki yazar tarafından puanlanmıştır. İki yazarın puanlayıcı güvenilirliği korelasyon katsayısıyla belirlenmiştir. Araştırmada makale (a), puanlayıcı (r) ve ölçek (i) olarak gösterilmek üzere genellenebilirlik kuramıyla değişkenlerin çapraz tasarlandığı  $a \times r \times i$  deseni analiz edilmiştir. Araştırma sonuçlarına göre yazarlar arasındaki uyum yüksek ve iki ölçeğinde eşit derecede iyi olduğu belirlenmiştir. İlk ölçeğin daha kısa olduğu böylelikle uygulamada daha az zaman aldığı, bununla birlikte ikinci ölçekten daha az bilgi sağladığı belirlenmiştir.

Wang (2005), çalışmasında geleneksel yöntem ve Genellenebilirlik Kuramı ile güvenilirlik hesaplayarak bir karşılaştırma yapmıştır. Bu amaçla öğrencilerin yazma becerilerini ölçmek için hazırlanmış sınav verileri kullanmıştır. Çalışmada öğrenci (p), madde (i) ve beceri (t) olmak üzere  $px(i:t)$  deseniyle G çalışması yapılmıştır. Ayrıca araştırmacı çalışmada  $\alpha$  güvenilirlik katsayısını hesaplamış ve bu değeri G katsayısıyla karşılaştırmıştır. Çalışmanın sonucunda  $\alpha$  ve G katsayısının birbirine yakın olduğu sonucuna ulaşılmıştır.

Gao ve Brennan (2001), performans görevleri üzerine yaptıkları çalışmada eğitim alanında yer alan çalışmaların ve bireylere verilen görevlerin hata varyansının en önemli kaynağı olduğu ve gözlemcilerin hata varyansına katkısının en az olduğunu ortaya koymuşlardır. Ayrıca performans görevleri güvenilirliğinin, içerik alanına

(dinleme ya da yazma) bağılı olduğu ortaya çıkmıştır. Çalışmada en dikkat çekici nokta, araştırmacıların çeşitli çalışmalar arasında varyans değerlerini ve güvenilirlik tahminlerini raporlaştırmalarıdır.

Bruckner, Yoder ve McWilliam (2006), G Kuramı kullanarak okul öncesi öğrencilerinden dil bilgisi ve fonolojik bozukluğu olan öğrencileri gözlemleyip ulaştırdıkları gözlem sonuçlarını değerlendirmişlerdir. Araştırmacılar araştırma deseninde farklı görev, öğrenci, puanlayıcı ve durumların etkileşimlerini incelemişlerdir. Araştırma sonuçlarına göre puanlayıcı genelinde öğrenci puanlarının güvenilir olduğu sonucuna ulaşılmış, ancak oturum-öğrenci etkileşiminde yani hata kaynağında oturum(durum) ile güvenilmezliğe neden olan sonuçlar arasında ilişki olduğu belirtilmiştir. Oturum veya durumun aslında çoğaltılmış bir görev olduğu çalışmanın sonuçlarına göre, görevin örneklenmesinde olası bir problem olabileceğinden, oturum veya durumun görev-öğrenci etkisine alternatif olarak yorumlanabileceği belirtilmiştir.

Wasserman, Levy ve Loken (2008), araştırmalarında psikoterapi sürecinin güvenilirlik ve geçerlik değerlerini G Kuramı ile değerlendirmişlerdir. Araştırmada farklı oturumlar, Borderline Kişilik Bozukluğu için Psikoterapi Süreci Değerlendirme Ölçeği'ne ait maddeler, ölçeğe göre değerlendirme yapan kodlayıcılar ve hasta hata kaynakları ve varyans değerleri araştırılmıştır. Ayrıca güvenilirlik katsayısının daha yüksek gözlemlerinin tahmin edilmesi istendiği için veriler üzerinden karar çalışması yapılmaya çalışılmış ve sonuçlar yorumlanmıştır. Yapılan karar çalışması sonucuna göre hastaların daha fazla oturumda, ölçülen ölçeğin madde sayısının fazla olması durumunda ve daha çok kodlayıcı yani gözlemleyici tarafından değerlendirilmesinin güvenilirlik katsayısını arttırabileceği sonucuna ulaşılmıştır.

Thiessen (2006), bir matematik yerleştirme sınavında öğrenciler hakkında doğru yerleştirme kararları verebilmek için testte ihtiyaç duyulan soru sayısını belirlemek amaçlı yaptığı çalışmasında Genellenabilirlik Kuramı'nı kullanmıştır. Çalışmaya katılan öğrencilerden 50'si fen, 51'i sosyal bilimler ve 57'si de alanlarına karar veremeyen öğrencilerden oluşmaktadır. Sınav cebir, geometri ve fonksiyon içerik kategorilerinden oluşan 30 sorudan oluşmaktadır. Araştırmacı desenin dengelenmemiş desen olmaması için farklı alanlardaki öğrenci sayılarını 50 olarak belirlemiştir. Araştırma sonuçlarına

göre yapılan karar çalışmasında yerleştirme testinin üç alanında da (cebir, geometri ve fonksiyonlar) en az sekiz maddenin olması gerektiği belirlenmiştir.

#### **2. 4. 2. Farklı Ölçme Araçları İle İlgili Araştırmalar**

Akdeniz (2003) çalışmasına, fizik dersine dayalı olarak geliştirilen kısa cevaplı ve çoktan seçmeli testlerden elde edilen madde ve test istatistiklerinin karşılaştırılmasını konu etmiştir. Araştırma 2001-2002 Eğitim- Öğretim Yılı'nda öğrenim gören 427 öğrenci üzerinde yapılmıştır. Uygulamada kullanılan testler; araştırmacı tarafından geliştirilmiş ve 29 maddeden oluşan biri çoktan seçmeli, diğeri kısa cevaplı iki sayısal başarı testidir. Kısa cevaplı test, çoktan seçmeli teste dayalı olarak oluşturulmuştur. Seçme gerektiren ve kısa cevap gerektiren maddelerden oluşturulan testlerin geçerlik katsayıları karne notları ölçüt alınarak Pearson Momentler Çarpımı Korelasyon Katsayıları ile, güvenirlik katsayıları ise KR 20 güvenirlik katsayısı ile bulunmuştur. Elde edilen analiz sonuçlarına göre, kısa cevaplı ve çoktan seçmeli maddelerden oluşan testlerin madde güçlük indeksleri arasında manidar farklılık olmadığı gözlenmiştir. Kısa cevaplı testin madde ayırıcılık gücü indeksleri, çoktan seçmeli testin madde ayırıcılık gücü indekslerinden daha yüksek bulunmuştur. Testlerin geçerliği, öğrencilerin birinci dönem karne notları ölçüt alınarak hesaplanmış ve kısa cevaplı testin geçerlik katsayısı, çoktan seçmeli teste göre daha yüksek bulunmuştur. Araştırma sonuçlarına göre, değişik madde tipinin kullanılmasının testlerin psiko-metrik özelliklerinden test güvenirlikleri üzerinde etkili olmadığı sonucuna ulaşılmıştır.

Özdemir (1997) araştırmasında, aynı davranışları ölçmek üzere geliştirilen kısa cevaplı, çoktan seçmeli ve kompozisyon tipi üç ayrı testin psiko-metrik özellikleri bakımından karşılaştırılmasını yapmıştır. Testlerin hazırlanmasında, belirlenen 17 davranışa ilişkin kısa cevaplı test için 44, çoktan seçmeli test için 41 ve kompozisyon tipi test için 14 madde hazırlanmış, ön deneme uygulaması sonucunda yapılan madde analizi yardımıyla, esas uygulamaya 17 davranış ölçmek üzere kısa cevaplı test için 17, çoktan seçmeli test için 17 ve kompozisyon tipi test için 11 madde seçilmiştir. Hazırlanan testler, 1996-1997 öğretim yılında Hacettepe Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü'nde okuyan ve İstatistik I dersini alan 173 öğrenciye uygulanmıştır. Testlere ilişkin madde analizlerinde Excel paket programı kullanılmış, testlerin karşılaştırılması için, madde ve test istatistiklerinin dağılımına ilişkin tablo ve

grafikler hazırlanmıştır. Kompozisyon tipi maddelerden oluşan testin madde güçlük indeksleri kısa cevaplı maddelerden oluşan testin madde güçlük indekslerine göre daha düşük, çoktan seçmeli maddelerden oluşan testin madde güçlük indeksleri ise her iki testin madde güçlük indekslerine göre daha yüksek olarak gözlenmiştir. Araştırma sonuçlarına göre; çoktan seçmeli maddelerden oluşan testin madde ayırıcılık gücü indekslerinin diğer iki teste göre daha yüksek olduğu, bu maddelerin ölçülmek istenilen davranışa sahip olanlarla olmayanları daha iyi ayırdığı, kompozisyon tipi maddelerden oluşan testin madde varyanslarının çoktan seçmeli ve kısa cevaplı maddelerden oluşan testlere göre daha yüksek olduğu; dolayısıyla bireyler arası farkları daha iyi gösterdiği, kısa cevaplı ve çoktan seçmeli maddelerden oluşan iki testin madde güvenilirlik indekslerinin birbirine daha yakın ve kompozisyon tipi maddelerden oluşan testin madde güvenilirlik indekslerinden daha yüksek olduğu bulunmuştur. Üç testin güvenilirlikleri karşılaştırıldığında, çoktan seçmeli testin diğer iki teste göre daha yüksek bir güvenilirliğe sahip olduğu sonucuna ulaşılmıştır.

Yalçın (2012), yapmış olduğu çalışmada; aynı hedef ve davranışlara yönelik olarak hazırlanmış, kimya eğitiminde kullanılan farklı ölçme türlerinden objektif ölçme türü olarak bilinen, çoktan seçmeli test, kısa cevaplı test ve doğru-yanlış tipi testleri çözelti konusu için karşılaştırmayı ve aralarında anlamlı bir ilişkinin olup olmadığını tespit etmeyi amaçlamıştır. Aynı hedef ve davranışlara yönelik üç farklı türdeki ölçme aracını, 2008-2009 Eğitim-Öğretim Yılı'nda 32 öğrenciye uygulamıştır. Araştırmada kullanılmak üzere, kısa cevaplı test, çoktan seçmeli test ve doğru-yanlış tipi test türünde aynı hedef davranışlara yönelik ve aynı sayıda soru maddesi içeren üç farklı test, araştırmacı tarafından geliştirilmiştir. Araştırmanın sonucunda; çözelti konusu için uygulanan çoktan seçmeli test ile kısa cevaplı test ve doğru-yanlış tipi test arasında anlamlı bir ilişki bulunmuştur. Uygulanan kısa cevaplı test ile doğru-yanlış tipi test arasında ise anlamlı bir ilişki bulunmamıştır. Ayrıca, çözelti konusu için uygulanan kısa cevaplı test, çoktan seçmeli test ve doğru-yanlış tipi test ile öğrencilerin okullarında almış oldukları kimya dersi başarı notları arasında da anlamlı bir ilişki bulunmamıştır.

Demir (2010), araştırmasında; PISA 2003 ve PISA 2006 uygulamaları kapsamında, Türkiye'de öğrencilerin bilişsel alan testlerinde yer alan soru tiplerinden hangilerinde başarı yüzdelerinin daha yüksek ve hangilerinde başarı yüzdelerinin daha düşük olduğunu belirlemek istemiştir. Araştırma örneklemini; Türkiye'de PISA 2003 ve

PISA 2006 uygulamalarına katılan 15 yaş grubu öğrencilerden oluşmaktadır. Yapılan analizler sonucunda ulaşılan sonuçlara göre, Türkiye'de öğrencilerin soru tiplerine göre tepkileri ,alt alanlara ve soru tiplerine göre farklılık göstermektedir. Okuma becerileri ve fen bilimleri okuryazarlığı alt alanlarında Türkiye'de öğrencilerin başarı yüzdelerinin en yüksek olduğu soru tipi `çoktan seçmeli' sorular, matematik okuryazarlığı ve problem çözme becerileri alt alanlarında ise bu durumun farklılık gösterdiği sonucuna ulaşılmıştır. Türkiye'de öğrencilerin, yapılandırılmış (çoktan seçmeli, karmaşık çoktan seçmeli ve yarı yapılandırılmış sorular) soru tiplerinde, yanıtları kendilerinin oluşturması beklenen diğer soru tiplerine (kısa cevaplı, açık uçlu) göre başarı düzeylerinin daha yüksek olduğu gözlenmiştir.

Tosyalı (2002), çalışmasında sayısal yetenek testlerinde çoktan seçmeli, kısa cevaplı ve doğru-yanlış testler kullanmanın, madde türlerindeki farklılıklardan dolayı madde ve test istatistikleri üzerinde nasıl bir etki gösterdiğini araştırmıştır. Araştırma, 2001-2002 Eğitim ve Öğretim Yılı'nda Ankara' da okuyan 133 öğrenci üzerinde yapılmıştır. Uygulamada kullanılan testler araştırmacı tarafından geliştirilmiş ve doğru-yanlış testler, şans başarısı etkisi test yönergesi ve puanlama biçimi değiştirilerek azaltıldığında ne gibi sonuçlarla karşılaşılacağına görülmesi amacıyla iki form oluşturulmuştur. Araştırma sonuçlarına göre, doğru- yanlış testlerinde şans başarısı etkisi test yönergesi ve puanlama biçimi değiştirilerek azaltıldığında, madde ayıricılık gücü indekslerinin en yüksek değerine ulaştığı ve hatta çoktan seçmeli ve kısa cevaplı testin maddeleri ile aynı şekilde değişiklik gösterdiği sonucuna ulaşılmıştır. Hipotez testi sonuçlarına göre çoktan seçmeli, kısa cevaplı ve 0-1 puanlamalı doğru-yanlış testlerinin geçerlikleri arasında fark olmadığı görülmüştür. Buna karşılık test geçerliği, doğru- yanlış testlerinde verilen önermenin doğruluğunun gerekçesi ile sorulması ve buna göre puanlanmasından olumlu yönde etkilenmiştir.

## **BÖLÜM III**

### **YÖNTEM**

Bu bölümde, araştırmanın türü, araştırmanın yapıldığı grup, veri toplama araçları ve verilerin analiz kısımları yer almaktadır.

#### **3. 1. Araştırma Modeli**

Betimsel araştırmalar ne ve nasıl sorularına sistematik olarak cevap vererek, olay ve durumların detaylı olarak betimlenmesi amacıyla yapılır (Başol, 2008). Araştırma, G Kuramı ile mevcut sistemde öğretmenlerin aynı konu hakkında kullandıkları ölçme araçları arasında ilişkiyi belirleme çalışması olduğundan betimsel bir araştırma niteliği taşımaktadır. Ayrıca araştırma, G Kuramı kullanılarak ölçme sonuçlarına ait parametreler elde edilmesi ve puanlayıcı, ölçme aracı ve farklı desenlerin karşılaştırılması açısından betimsel nitelik taşımaktadır.

#### **3. 2. Evren ve Örneklem**

Araştırmanın çalışma evrenini 2013-2014 Eğitim-Öğretim yılında Kars il merkezinde öğrenim gören 8. sınıf öğrencileri oluşturmaktadır. Araştırma örneklemini ise Kars il merkezinde yer alan Atatürk Ortaokulu'nda öğrenim gören 48 öğrenci oluşturmaktadır. Araştırmacı uygulama kolaylığından dolayı amaçlı örnekleme gitmiştir.

Genellenebilirlik çalışmaları için çok büyük örneklere gerek duyulmadığına yapılan pek çok araştırma örneklemlerinden yola çıkılarak ulaşılmıştır (Kaya, 2011; Nalbantoğlu, 2009; Deliceoğlu ve Çıkrıkçı Demirtaşlı, 2012; Yelboğa, 2008; Taşdelen Teker, Eroğlu ve Baytemir, 2012; Güler ve Teşdelen Teker, 2012).

Araştırmanın yapıldığı okul devlet okuludur. Bu yüzden araştırma yapılabilmesi için gerekli izinler Gaziosmanpaşa Üniversitesi ve Kars İl Milli Eğitim Müdürlüğünden alınmıştır. Araştırma için uygulama süreci araştırmacının kendisi tarafından takip edilmiş ve öğrencilerin araştırma konusunda bilgilendirilmesine ve testleri ciddiye alarak cevaplamalarına katkıda bulunulmaya gayret edilmiştir.



### 3. 3. Araştırma Verileri

Araştırma için gerekli veriler, araştırmacı tarafından hazırlanan Klasik Sınav ve ölçme sürecinde daha önce kullanılmış olan sorular arasından seçilen Çoktan Seçmeli sınav sorularına verilen cevaplardan elde edilmiştir. Uygulanan ölçme araçları Ek 1 ve Ek 2’ de sunulmuştur. Çalışmada çoktan seçmeli test ve klasik sınav için ‘Hücre Bölünmesi ve Kalıtım’ ünitesi seçilmiştir. Bu ünitenin seçilme sebebi öğrencilerin hazırbulunuşluk seviyelerinin yakalanmak istenmesidir. Araştırma soruları ilgili konu alanına mensup üç uzman kanısına dayanılarak belirtke tablosu da kullanılarak kapsam ve yapı geçerliliği sağlanmaya çalışılmıştır.

Araştırma soruları için üniteye ait 20 kazanım ele alınmış olup öğrenci seviyesi de düşünülerek çoktan seçmeli test için ilk olarak 40 madde seçilmiştir. Hazırlanan bu sınav öncelikle iki konu alanı uzmanı ve bir dil uzmanına danışılarak deneme formatı için hazır hale getirilmiştir. Deneme uygulaması Kars il merkezinde yer alan Atatürk Ortaokulu, Fevzi Paşa Ortaokulu ve Halit Paşa Ortaokulu’nda öğrenim gören 96 öğrenci üzerinde yapılmış ve büyük ölçüde eksik olduğu belirlenen altı katılımcının cevapları çıkarılmıştır. Geriye kalan 90 kişinin cevapları dikkate alınmış ve deneme uygulamasının yapıldığı 90 kişiden oluşan grup nihai uygulamaya dahil edilmemiştir.

Tablo 1. Deneme Uygulamasından Elde Edilen Madde İstatistikleri

<b>Madde No</b>	<b>Madde Güçlük İndeksi (pj)</b>	<b>Madde Ayırt Edicilik İndeksi (rj)</b>
<b>1*</b>	.458	.333
2	.562	.625
<b>3*</b>	.646	.708
<b>4*</b>	.479	.542
<b>5*</b>	.500	.417
6	.417	.167
7	.625	.500
8	.542	.500
<b>9*</b>	.354	.375
<b>10*</b>	.458	.583

---

11	.562	.625
<b>12*</b>	.625	.667
13	.542	.833
<b>14*</b>	.688	.458
15	.604	.458
16	.042	-.083
<b>17*</b>	.625	.667
<b>18*</b>	.667	.417
19	.354	.375
20	.646	.625
<b>21*</b>	.625	.667
22	.479	.708
<b>23*</b>	.646	.458
<b>24*</b>	.604	.458
25	.542	.750
<b>26*</b>	.500	.500
<b>27*</b>	.646	.625
28	.375	.167
<b>29*</b>	.542	.833
<b>30*</b>	.458	.583
<b>31*</b>	.375	.667
32	.479	.792
<b>33*</b>	.500	.500
34	.458	.583
35	.524	.375
36	.542	.583
37	.479	.708
<b>38*</b>	.604	.708
<b>39*</b>	.562	.625
<b>40*</b>	.583	.667

---

Deneme uygulaması için test ve madde istatistikleri TAP.exe uygulaması kullanılarak elde edilmiş, Tablo 1'de her bir test maddesine ait madde güçlük ve ayırtedicilik katsayıları verilmiştir. Her bir kazanımın iki katı soru hazırlandığı için

madde güçlük ve madde ayırt edicilik gücü bakımından en iyi sonucu vermiş olan sorular seçilerek her biri dört seçenekli 22 adet soru nihai test uygulaması için seçilmiş ve nihai test formu belirlenmiştir.

Çoktan seçmeli test için yapılan ön uygulamanın betimsel istatistiklerine bakıldığında testin aritmetik ortalamasının 50.16, testten alınan minimum puanın 15 ve maksimum puanın 92 olduğu gözlenmiştir. Aşağıda verilen Tablo 2’ de çoktan seçmeli testin ön uygulamasına ait betimsel istatistiklere yer verilmiştir.

Tablo 2. Çoktan Seçmeli Testin Ön Uygulamasına Ait Betimsel İstatistikler

Öğrenci Sayısı (N)	90
Madde Sayısı (K)	40
Aritmetik Ortalama (X)	50.16
Varyans ( s <sup>2</sup> )	468.85
Standart Sapma (s)	21.65
En Düşük Puan ( Min.)	15.00
En Büyük Puan (Max.)	92.00
Ortalama Güçlük	.523
Ortalama Ayırt Edicilik	.544

**p < .01**

Bir maddenin güçlük katsayısı 0.00 ile +1.00 arasında değişir. Hiç kimsenin doğru cevaplayamadığında maddenin güçlük katsayısı sıfır; bütün öğrencilerin doğru cevapladığı bir maddenin güçlük katsayısı ise +1.00'dir. Madde ayırt edicilik gücü katsayısı ise -1.00 ile +1.00 arasında değişmektedir. Madde ayırt edicilik gücü katsayısı .40 ve üzerinde olan maddeler çok iyi maddelerdir (Başol, 2014).

Çoktan seçmeli testin ön uygulamasından elde edilen madde istatistiklerine göre hazırlanan klasik sınav sorularının doğrultusunda 22 madde nihai uygulama için seçilmiştir. Madde istatistikleri ve konu alanına göre uygun bulunan maddeler Tablo 1 de koyu renk ile belirtilmiştir. Çoktan seçmeli test için seçilen maddelerin, madde güçlükleri farklı değerde oldukları için KR 20 güvenilirlik değeri hesaplanmış ve bu katsayının .896 olduğu görülmüştür. KR 20 ile hesaplanan güvenilirlik testin kendi içinde tutarlılığının bir ölçüsü olup bu değer yüksek çıkması test maddelerinin kendi içinde

tutarlı olduğunu ve güvenilir bir test olduğunu ortaya çıkarmaktadır (Başol, 2014).

Yazılı sınav için iki konu alanı uzmanının görüşüne başvurularak kapsam geçerliliğinin sağlanması amacıyla sorular hazırlanmış ve bir dil uzmanına danışılarak uygulama formu hazır hale getirilmiştir. Soruların yanlış anlaşılmalara neden olmaması ve tarafsızlığa hizmet etmesi açısından, bir kız ve bir erkek öğrenciye önceden çözdürülmüştür. Sınavın uygulandığı bu iki öğrenci için uygulanan sınav sonrası öğrenci görüşleri ele alındığında cinsiyete göre taraf oluşturacak bir durumun olmadığı sonucuna ulaşılmıştır. Ayrıca bu iki öğrenci nihai uygulama grubu arasında yer almamıştır.

Çoktan seçmeli test ve Yazılı (Klasik) sınav Kars İl Merkezinde yer alan Atatürk Ortaokulunda öğrenim gören 48 katılımcıya birer hafta ile uygulanmış ve uygulamalar araştırmacı tarafından bireysel olarak gözlemlenmiştir.

### **3. 4. Verilerin Analizi**

Araştırmacı tarafından geliştirilen ölçme araçlarından elde edilen verilerin analizinde TAP.exe, SPSS 16.0 ve G kuramı analizleri için EduG paket programları kullanılmıştır.

Deneme uygulaması sonuçlarına göre madde seçiminde TAP.exe, çoktan seçmeli test ve klasik sınav analizleri için SPSS 16.0 ve Genellenebilirlik Kuramı çalışmasının analizleri için EduG paket programları kullanılmıştır.

İlk olarak belirlenen ölçme araçları ile gerekli uygulamalar yapılmış olup çoktan seçmeli nihai test maddeleri ortak sonuçlar doğuracağından tek bir puanlayıcı tarafından hazırlanmış olan yazılı sınav ise üç farklı puanlayıcı tarafından puanlanmıştır. Puanlayıcılara araştırmacı tarafından puanlama cetveli verilmiş ve puanlama için gerekli süre sağlanmıştır. Puanlayıcılar birbirlerini tanımamakta, farklı özelliklerdeki okullarda görev yapmakta ve farklı mesleki kıdemlere sahiptirler

## BÖLÜM IV

### BULGULAR

Bu bölümde araştırmadan elde edilen verilerin araştırma sorularına ilişkin bulgulara yer verilmiştir. Araştırma sorularına geçilmeden önce betimsel istatistiklere yer verilmiştir.

#### 4. 1. Alt Problem 1; Performansın Ölçülmesinde Kullanılan Çoktan Seçmeli Sınavın İncelenmesi

Öğrencilere ilk olarak uygulanan çoktan seçmeli sınav kazanımlar dahilinde 22 soru üzerinden değerlendirilmiştir. Bunun için öncelikle öğrencilerin doğru cevapları hesaplanmış bunlar 100 puan üzerinden dönüşüm yapılmış olup, öğrencilere yapılan çoktan seçmeli sınava ait istatistikler Tablo 3' de verilmiştir.

Tablo 3. Çoktan Seçmeli Test İle Yapılan Nihai Uygulamaya Ait Betimsel İstatistikler

Soru									
Sayı	n	Ortalama	Medyan	Mod	Mak.	Min.	Ranj	Çarpıklık	Basıklık
22	48	68.37	68.18	68.18	100.00	18.18	81.82	-.55	-.33

**p < .01**

Uygulanan çoktan seçmeli testte her bir maddeden alınabilecek en yüksek puan bir, testten alınabilecek en yüksek puan 22' dir. Alınan puanlar 100 üzerinden değerlendirmeye alınmış ve istatistiksel işlemler bu puanlar üzerinden yapılmıştır. Dönüştürülen puanlara göre çoktan seçmeli testin ortalaması 68.37, medyanı 68.18, modu 68.18' dir. Testten alınan standart puan en yüksek puan 22 sorunun hepsini doğru cevaplayan üç kişi için 100, testten alınan en düşük puan ise dört doğru ile 18.18 olarak hesaplanmıştır. Puan serisinin çarpıklık katsayısı -.55 ve basıklık katsayısı -.33 olarak hesaplanmıştır.

Çarpıklık katsayısı -1 ile 1 arasında olan veriler ile basıklık katsayısı -1 ile 1 arasında olan veriler normal dağılım gösteren verilerdir (Büyüköztürk, 2006; Kalaycı, 2008). Çarpıklık katsayısının (-.55) değerinin -1 ile 1 arasında hesaplandığı ve basıklık değerinin de (-.33) -1 ile 1 arasında değer aldığı için çoktan seçmeli testten alınan

notların normal dağılım gösterdiği kabul edilmiştir. Çarpıklık değerinin negatif çıkması çoktan seçmeli testten elde edilen notların dağılımının hafif sola çarpık olduğu ve basıklık değerinin negatif olması da normalden daha basık bir dağılım olduğunu göstermektedir.

Çoktan seçmeli test için hesaplanan standart sapma değeri 4.6309 olarak hesaplanmıştır. Bu değerden yola çıkılarak çoktan seçmeli teste ait varyans değeri 21.44 olarak ortaya çıkmıştır.

#### **4. 1. 1. Performansın Ölçülmesine Yönelik G Kuramına Göre Kestirilen Parametrelerin Varyansları ve Toplam Varyansları Açıklama Yüzdeleri**

Çoktan seçmeli test için birey (b) ve madde (m) değişkenlerinin değişimlerini ve varyans kaynaklarının oranlarını belirlemek için tek değişkenli G (Genellenebilirlik) çalışması yapılmıştır.

Tablo 4. Tek Değişkenli G Çalışması Sonucunda Ölçmenin Kestirilen Varyansları ve Toplam Varyansı Açıklama Oranları

<b>Varyans Kaynağı</b>	<b>d</b>	<b>Toplam Kareler</b>	<b>Kareler Ortalaması</b>	<b>Varyans</b>	<b>%</b>
b	47	45.81439	.97477	.037	16.9
m	21	20.23485	.96356	.017	7.6
bm	87	162.31061	.16445	.164	75.5
Toplam					100

Tablo 4 incelendiğinde birey (b) ana etkisi için kestirilen varyans bileşeninin (.037) toplam varyansın % 16.9' unu açıkladığı görülmektedir. Tek değişkenli modelle yapılan incelemede bireyler için kestirilen varyans bileşeni, toplam varyans içinde en yüksek ikinci paya sahip olan varyans bileşenidir. Genellenebilirlik çalışmalarında, birey ana etkisi evren puanı varyansı olarak değerlendirilir ve ölçülen özellik açısından bireyler arası farklılaşmayı ifade eder (Shavelson & Webb, 1991; Brennan, 2001).

Bireyler için kestirilen varyansın toplam varyans içindeki oranının en büyük olması istenilen bir durumdur. Bu durum, ölçme ile elde edilen boyutta bireyler arası farklılıkların ortaya çıkarılabildiğinin bir göstergesidir (Güler ,2008).

Madde (m) ana etkisi için tek deęişkenli modelle yapılan G çalışmasında kestirilen varyans bileşeni ( .017) toplam varyansın % 7.6' sını açıklamaktadır. Madde ana etkisinin varyans bileşeni büyüklüğün, toplam varyans deęişkeni büyüklüğünde üçüncü ve en az orana sahiptir.

Birey x madde ortak etkisi (.164) toplam varyansın % 75.5' ini açıklamaktadır. Birey x madde ortak etkisi tek deęişkenli modelle yapılan G çalışmasında elde edilen en büyük varyans deęeridir. Bu durum; bu ölçme için birey x madde ortak etkisinden kaynaklanan farklılığın büyük olduğunu, belli bireylerin baęıl durumlarının bir maddeden dięerine çok farklılaştığını göstermektedir. Ayrıca birey x madde varyans deęerinin büyük olması birey ve madde ortak etkisi veya tesadüfü hataların büyük olabileceęi anlamına geliyor olabilir.

#### **4. 1. 2. Performansın Ölçülmesine Yönelik Madde Sayılarının Arttırılıp Azaltılması Senaryolarına Göre Hesaplanan Hata Varyansları**

Yapılan çoktan seçmeli test için 22 madde ve bu madde sayısının azaltılıp arttırılması durumunda hata varyanslarının deęişimlerini belirlemek amacıyla G kuramı çalışmasıyla yapılan K çalışmasına göre hesaplanan hata varyansları deęişimleri Tablo 5' de verilmiştir.

Tablo 5. Performansın Ölçülmesine İlişkin Yapılan K Çalışması İle Madde Sayıları Senaryolarına Göre Baęıl ve Mutlak Hata Varyansları

Madde Sayısı	Baęıl	Mutlak
18	.00914	.01006
20	.00822	.00905
22	<b>.00747</b>	<b>.00823</b>
24	.00685	.00755
26	.00632	.00697

Tablo 5' de çoktan seçmeli test için madde sayılarının arttırılıp azaltılması durumlarına baęlı olarak hesaplanan baęıl ve mutlak hata varyansları verilmiştir. Buna göre, çoktan seçmeli 22 maddelik nihai testten elde edilen baęıl hata varyansı .00747 ve mutlak hata varyansı .00823 olarak kestirilmiştir. Bu deęerler incelendiğinde, asıl uygulamadan elde edilen mutlak hata varyansı, baęıl hata varyansından yüksektir. Madde sayılarının arttırılıp azaltılması ile yapılan senaryolara göre baęıl ve mutlak hata

varyansları hesaplanmıştır. Madde sayısının dört azaltılması durumunda (18 madde üzerinden) yapılan hesaplamada bağıl hata varyansı .00914 ve mutlak hata varyansı .01006 olarak kestirilmiştir. Bu senaryo çalışması üzerinden elde edilen hata varyanslarında; mutlak hata varyansının bağıl hata varyansından yüksek olduğu gözlenmiştir. Madde sayısının iki azaltılması durumunda (20 madde üzerinden) yapılan hesaplamalarda bağıl hata varyansı .00822 ve mutlak hata varyansı .00905 olarak kestirilmiştir. Madde sayısı azaldıkça, bağıl hata varyansı mutlak hata varyansına yaklaşmıştır. Bu senaryo çalışması üzerinden elde edilen hata varyanslarında; mutlak hata varyansının bağıl hata varyansından yüksek olduğu gözlenmiştir.

Madde sayısının iki artırılması durumunda (24 madde üzerinden) yapılan hesaplamalarda bağıl hata varyansı .00685 ve mutlak hata varyansı .00755 olarak kestirilmiştir. . Bu senaryo çalışması üzerinden elde edilen hata varyanslarında yine; mutlak hata varyansının bağıl hata varyansından yüksek olduğu gözlenmiştir. Son olarak madde sayısının dört artırılması durumunda (26 madde üzerinden) yapılan hesaplamalarda bağıl hata varyansı .00632 ve mutlak hata varyansı .00697 olarak kestirilmiştir. Bu senaryo çalışmasına göre yine mutlak hata varyansının bağıl hata varyansına göre biraz daha yüksek çıktığı gözlenmiştir.

Tablo 5' e göre çoktan seçmeli test maddelerinin artırılıp azaltılmasına göre bağıl ve mutlak hata varyanslarının değişim gösterdiği gözlenmiştir. Madde sayılarının artırılması ile hem bağıl hem de mutlak hata varyanslarının azalma gösterdiği ortaya çıkmıştır. Ayrıca kestirilen hata varyanslarından mutlak hata varyansının, bağıl hata varyansına göre yapılan tüm senaryo denemelerinde daha yüksek değere sahip olduğu gözlenmiştir. Bu sonuçlara göre madde sayısı arttıkça tahminin güvenilirliğinin artacağı sonucuna varılabilir.

#### **4. 1. 3. Performansın Ölçülmesine Yönelik Madde Sayılarının Arttırılıp Azaltılması Senaryolarına Göre Hesaplanan G ve $\Phi$ Katsayıları**

Performansın ölçülmesinde kullanılan çoktan seçmeli test için 22 madde ve madde sayısının arttırılıp azaltılması durumlarında G Kuramı çalışması ile yapılan K çalışması sonucu elde edilen G ve  $\Phi$  katsayıları Tablo 6' da verilmiştir.



Tablo 6. Performansın Ölçülmesine İlişkin Yapılan K çalışması İle Madde Sayıları Senaryolarına Göre G ve Phi Katsayıları

Madde Sayısı	$\Phi$	G
18	.80126	.78545
20	.81750	.80267
22	<b>.83130</b>	<b>.81734</b>
24	.84315	.82997
26	.85345	.84097

Tablo 6' da görüldüğü gibi çoktan seçmeli testin madde sayılarının arttırılıp azaltılması durumlarına göre hesaplanan G ve  $\Phi$  katsayıları verilmiştir. Tabloya göre, madde sayısının nihai testteki değerine göre yapılan analiz sonuçlarına göre;  $\Phi$  katsayısı .83130 ve G katsayısı .81734 olarak, madde sayısının dört azaltılması durumunda (18 madde üzerinden)  $\Phi$  katsayısı .80126 ve G katsayısı .78545 olarak, madde sayısının iki azaltılması durumunda (20 madde üzerinden)  $\Phi$  katsayısı .81720 ve G katsayısı .80267 olarak kestirilmiştir.

Madde sayılarının arttırılması üzerinden yapılan analiz sonuçlarına göre; madde sayısının 2 arttırılması durumunda (24 madde üzerinden)  $\Phi$  katsayısı .84315 ve G katsayısı .82997 olarak, madde sayısının 4 arttırılması durumunda (26 madde üzerinden)  $\Phi$  katsayısı .85345 ve G katsayısı .84097 olarak kestirilmiştir.

Tablo 6 incelendiğinde, madde sayısının azaltılması durumlarında  $\Phi$  katsayısı ve G katsayılarının azaldığı, madde sayısının arttırıldığı durumlarda  $\Phi$  katsayısı ve G katsayılarının arttığı gözlemlenmiştir.

#### 4. 2. Alt Problem 1; Performansın Ölçülmesinde Kullanılan Klasik Sınavın İncelenmesi

Performansın ölçülmesine yönelik uygulanan klasik sınav 11 maddeden oluşmaktadır. Uygulanan sınav üç farklı puanlayıcı tarafından puanlanmış ve puanlayıcılar üzerinden elde edilen verilerle işlemler gerçekleştirilmiştir. Klasik sınava yönelik puanlayıcılardan elde edilen puanlara ait betimsel istatistikler Tablo 7' de verilmiştir.

Tablo 7. Performansın Ölçülmesinde Yapılan Klasik Sınav İçin 3 Puanlayıcı Ait Betimsel İstatistikler (N=48)

İstatistikler	1. Puanlayıcı	2. Puanlayıcı	3. Puanlayıcı
Ortalama	56.1875	45.4792	34.7083
Medyan	56.500	44.500	29.500
Mod	34	43	30
Std. Sapma	2.36208	2.35833	2.4114
Varyans	557.943	556.170	581.360
Çarpıklık	-.184	.248	.609
Basıklık	-.970	-.990	-.769
Minimum	8	8	2
Maksimum	96	83	84
$\alpha$ güvenirligi	.85	.87	.87

$p < .01$

Tablo 7 incelendiğinde, 48 öğrencinin 11 madde üzerinden aldıkları puanlara ilişkin en yüksek ortalama Birinci puanlayıcıya ait ve 56.1875 dir. En düşük ortalama ise 34.7083 ile üçüncü puanlayıcıya aittir. İkinci puanlayıcı 45.4792 ile puanlayıcı ait ortalama değeri ise bu iki değer arasında yer almaktadır. Birinci puanlayıcıya ilişkin ortanca değer aritmetik ortalama dan yüksektir ve puanlayıcılara ait puanların sola çarpık dağılım gösterdiği gözlenmiştir. İkinci ve üçüncü puanlayıcıya ilişkin ortanca değerlerinin aritmetik ortalamadan küçük olması ise puanlayıcıya ait puanların sağa çarpık dağılım gösterdiğini ortaya koymaktadır. Bu durum çarpıklık katsayılarının birinci puanlayıcıya ait puan değerleri için negatif, ikinci ve üçüncü puanlayıcılara ait puan değerleri için pozitif çıkmasıyla da görülmektedir.

Basıklık katsayısına bakıldığında; her üç puanlayıcıya ait puan değerlerinin 0' dan küçük çıkması puanların normalden daha basık dağılım gösterdiğini ortaya koymaktadır. Puanlayıcıların verdikleri puan değerlerine ait Cronbach Alfa ( $\alpha$ ) güvenirlilik katsayıları birbirine yakın ve yüksek değerlerdir.

Puanlayıcıların 11 madde üzerinden verdikleri puanlar arasındaki korelasyon değerleri Tablo 8' de verilmiştir.

Tablo 8. Puanlayıcıların 11 Maddeye Verdikleri Puanlar Arasındaki Korelasyon Katsayıları

	1. Puanlayıcı	2. Puanlayıcı	3. Puanlayıcı
1. Puanlayıcı	-	0.93	0.90
2. Puanlayıcı		-	0.97

**p < .01**

Tablo 8 incelendiğinde, puanlayıcıların 11 madde üzerinden verdikleri puanların birbirleri ile olan korelasyon katsayılarının oldukça yüksek olduğu gözlenmektedir. Elde edilen bu değerlere göre puanlayıcılar arasındaki uyumun oldukça yüksek olduğundan söz edilebilir.

#### 4. 2. 1. Performansın Ölçülmesine Yönelik G Kuramına Göre Kestirilen Parametrelerin Varyansları ve Toplam Varyansları Açıklama Yüzdeleri

Matematik performansının ölçülmesine yönelik hazırlanan 11 maddelik klasik ölçme aracının G çalışması ile elde edilen varyanslarını ve varyans yüzdelerini hesaplamak için tümüyle çaprazlanmış b x m x p modeli uygulanmıştır. Ölçmenin uygulandığı 48 öğrenci, 11 madde ve üç puanlayıcıdan oluşan verilerde tek değişkenli modelle yapılan G çalışması için; kestirilen varyans bileşenleri ve toplam varyansı açıklama yüzdeleri b, m ve p ana etkileri ile bm, bp, mp, ve bmp ortak etkileri Tablo 9’ da verilmiştir.

Tablo 9. Tek Değişkenli G Çalışması Sonucunda Ölçmenin Kestirilen Varyansları ve Toplam Varyansı Açıklama Oranları

Varyans Kaynağı	Sd	Toplam Kareler	Kareler Ortalaması	Varyans	%
b	47	5193.65657	110.50333	3.08173	16.0
m	10	44.32576	4.43258	-.02686	.0
p	2	127.97096	63.98548	.13021	.7
bm	470	8652.70455	18.41001	1.15344	6.0
bp	94	502.51389	5.34589	-.87307	.0
mp	20	96.80682	4.84034	-.21061	.0
bmp	940	14052.70833	14.94969	14.94969	77.4
Toplam					100

Tablo 9 incelendiğinde, birey (b) ana etkisi için kestirilen varyans bileşenini (3.08173) toplam varyansın % 16' sini açıklamaktadır. Tek değişkenli modelle bireyler için kestirilen varyans bileşeni, toplam varyans içinde en yüksek ikinci sırada paya sahiptir.

Madde (m) ana etkisi için kestirilen varyans bileşeni tek değişkenli modelle yapılan G çalışmasına göre kestirilen varyans bileşeni eksi değer aldığı için (-.02686) toplam varyansı açıklama yüzdesi içinde (%0) bir etkiye sahip olmadığı görülmüştür. Varyansın *sifir* alınmasının nedeni G Kuramı çalışmalarında varyans değerinin negatif çıkması durumlarında uygulanan üç farklı yöntemden biri olmasıdır (Brennan, 2001). Shavelson ve Webb (1981)'e göre negatif varyanslar örnekleme hatalarından ya da yanlış model seçiminden kaynaklanmış olabilir.

Shavelson ve Webb (2005) negatif varyans söz konusu olduğunda dört çözüm önerisi olduğunu belirtmiştir: Cronbach, Gleser, Nanda ve Rajaratnam (1972) negatif varyans değerinin yerine sıfır yazmayı önermişler, ikinci öneri olarak (Brennan, 2001) negatif varyansların sıfır alınmasını ancak beklenen ortalama kareler eşitliğinde negatif varyansların olduğu gibi kullanılmasını, üçüncü öneri ise (Shavelson ve Webb) Bayesian metod kullanılarak tahmin edilen varyans için en küçük değerini sıfır olarak değiştirilmesini, son olarak Searle (1987) maksimum olabilirlik modeli kullanılarak negatif varyansların önüne geçilmesini önermiştir (Akt. Shavelson ve Webb, 2005).

Puanlayıcı ana etkisinin G çalışması ile kestirilen varyans bileşeni (.13021) toplam varyansın % 0.7' ini açıklayarak toplam varyans içinde dördüncü sırada yer almaktadır. Puanlayıcı etkisinin tek değişkenli modelle yapılan G çalışması ile kestirilen varyans oranının düşük olması, puanlayıcıların tüm bireyler için yaptıkları puanlamalar arasında bir fark bulunmadığını, puanlamalar arasında da bir tutarlılığın olduğunu göstermektedir.

Birey x madde (bm) ortak etkisi (1.15344) toplam varyansın % 6' sını açıklamaktadır. Birey x madde ortak etkisi tek değişkenli modelle kestirilen en yüksek üçüncü değere sahip varyans değeridir. Bu da birey x madde ortak etkisinden kaynaklanan farklılığın büyük olduğunu, belli bireylerin bağlı durumlarının bir maddeden diğer maddeye çok farklılaştığını göstermektedir (Güler, 2008).

Birey x puanlayıcı (bp) ortak etkisi (-.87307) toplam varyansın % 0' ını açıklamaktadır. Madde x puanlayıcı (mp) ortak etkisi (-.21061) 0' ın altında değer aldığı için toplam varyans içerisinde açıklama yüzdesi %0 dır. Madde x puanlayıcı etkisinin tek değişkenli modele göre madde x puanlayıcı ortak etkisinden kaynaklanan bir farklılığın olmadığı yorumu yapılabilir.

Birey x madde x puanlayıcı (artık) ortak etkisi varyans bileşenide (14.94969) toplam varyansın % 77.4' ünü açıklamaktadır. Bu oran varyans değerleri arasından en büyük değerdir. Birey x madde x puanlayıcı (artık) varyansın büyük olması; birey, madde ve puanlayıcı ortak etkisi veya tesadüfi hataların büyük olabileceğinin bir göstergesi olabilir.

#### 4. 2. 2. Performansın Ölçülmesine Yönelik Madde Sayılarının Arttırılıp Azaltılması Senaryolarına Göre Hesaplanan G ve $\Phi$ Katsayıları

Uygulanan klasik sınava ait veriler üzerinden madde sayısı ve puanlayıcı sayılarının arttırılıp azaltılması durumlarına göre G kuramı kullanılarak K çalışması yapılmıştır. Yapılan K çalışmasına ait G ve  $\Phi$  katsayılarının değişimi Tablo 10' da verilmiştir.

Tablo 10. Performansın Ölçülmesine İlişkin Yapılan K çalışması ile Madde ve Puanlayıcı Sayıları Senaryolarına Göre Phi ve G Katsayıları

Madde Sayıları	Puanlayıcı Sayıları							
	2		3		4		5	
	G	$\Phi$	G	$\Phi$	G	$\Phi$	G	$\Phi$
9	.76272	.75063	.81883	.80949	.85010	.84253	.87003	.86368
11	.79711	.78391	<b>.84672</b>	<b>.83674</b>	.87391	.86592	.89109	.88443
13	.82279	.80874	.86717	.85671	.89120	.88289	.90627	.89938
15	.84271	.82797	.88280	.87196	.90432	.89576	.91774	.91068
18	.86539	.84986	.90039	.88912	.91897	.91014	.93050	.92324

Tablo 10 incelendiğinde; tek değişkenli modelle yapılan ölçme sonuçlarına göre 11 madde ve üç puanlayıcıya göre kestirilen G katsayısı .84672 ve  $\Phi$  katsayısı da .83674 olarak kestirilmiştir. Kestirilen katsayı değerlerine bakılarak G katsayısının  $\Phi$  katsayısından daha yüksek olduğu görülmektedir. Madde sayısının iki azaltılması ile

dokuz madde üzerinden kestirilen katsayılar incelendiğinde; puanlayıcı sayısının iki olması durumunda G katsayısı .76272 ve  $\Phi$  katsayısı .75063 olarak ortaya çıkmış, puanlayıcı sayısının üç olması durumunda ise G katsayısının .81883 ve  $\Phi$  katsayısı .80949 olarak, puanlayıcı sayısının dört olması durumunda G katsayısı .85010 ve  $\Phi$  katsayısı .84253 olarak ve son olarak puanlayıcı sayısının beş olması durumunda G katsayısının .87003 ve  $\Phi$  katsayısının .86368 olduğu ve G ve Phi katsayılarındaki artışın puanlayıcı sayısının ikiden üçe çıkması halinde, dörtten beşe çıkması durumuna göre daha fazla artış gösterdiği görülüyor.

Madde sayısının esas uygulama sayısı olarak 11 olarak sabit kalması ve puanlayıcı sayısının iki olması durumunda G katsayısı .79711 ve  $\Phi$  katsayısı .78391 olarak kestirilmiştir. Puanlayıcı sayısının dört olması durumunda ise G katsayısı .87391 ve  $\Phi$  katsayısı .86592, puanlayıcı sayısının beş olması durumunda G katsayısı .89109 ve  $\Phi$  katsayısı .88443 olarak kestirilmiştir.

Madde sayısının iki arttırılıp 13 olması; puanlayıcı sayısının iki olması durumunda G katsayısı .82279 ve  $\Phi$  katsayısı .80874, puanlayıcı sayısının üç olması durumunda G katsayısı .86717 ve  $\Phi$  katsayısı .85671 olarak, puanlayıcı sayısının arttırılıp dört olması durumunda G katsayısı .89120 ve  $\Phi$  katsayısı .88289, puanlayıcı sayısının beş olması durumunda G katsayısı .90627 ve  $\Phi$  katsayısı .89938 olarak kestirilmiştir.

Madde sayısının dört arttırılıp 15 olması halinde; puanlayıcı sayısının iki olması durumunda G katsayısı .84271 ve  $\Phi$  katsayısı .82797, puanlayıcı sayısının üç olması durumunda G katsayısı .88280 ve  $\Phi$  katsayısı .87196 olarak, puanlayıcı sayısının arttırılıp dört olması durumunda G katsayısı .90432 ve  $\Phi$  katsayısı .89576 ve puanlayıcı sayısının beş olması durumunda G katsayısı .91774 ve  $\Phi$  katsayısı .91068 olarak kestirilmiştir.

Madde sayısının yedi arttırılıp 18 olması ve; puanlayıcı sayısının iki olması durumunda G katsayısı .86539 ve  $\Phi$  katsayısı .84986, puanlayıcı sayısının üç olması durumunda G katsayısı .90039 ve  $\Phi$  katsayısı .88912 olarak, puanlayıcı sayısının arttırılıp dört olması durumunda G katsayısı .91897 ve  $\Phi$  katsayısı .91014 ve son olarak

puanlayıcı sayısının beş olması durumunda G katsayısı .93050 ve  $\Phi$  katsayısı .92324 olarak kestirilmiştir.

Tablo 10' a göre gerek bağıl değerlendirme durumlarında kullanılan G katsayısı ve gerek mutlak değerlendirme durumlarında kullanılan  $\Phi$  katsayılarının madde sayılarının ve puanlayıcı sayılarının artması durumunda yükseldiği ortaya çıkmıştır. Tüm madde ve puanlayıcı senaryolarında G katsayıları,  $\Phi$  katsayılarından yüksek değerlerde çıkmıştır. Madde sayısının aynı kalması durumunda puanlayıcı sayısının artması senaryolarında ortaya çıkan G ve  $\Phi$  katsayıları; puanlayıcı sayılarının aynı kalması durumunda madde sayısının arttırılması ile kestirilen G ve  $\Phi$  katsayılarına göre daha yüksek değerlerde ortaya çıkmıştır.

#### 4. 2. 3. Performansın Ölçülmesine Yönelik Madde sayılarının Arttırılıp Azaltılması Senaryolarına Göre Hesaplanan Mutlak ve Bağıl Hata Varyansları

Matematik performansının ölçülmesi için yapılan 11 madde ve üç puanlayıcı sayısına, madde ve puanlayıcı sayılarının arttırılıp azaltılmasına göre yapılan K çalışması sonucu kestirilen mutlak ve bağıl hata varyansları Tablo 11'de verilmiştir.

Tablo 11. Performansın Ölçülmesine İlişkin Yapılan K Çalışması İle Madde ve Puanlayıcı Sayıları Senaryolarına Göre Bağıl ve Mutlak Hata Varyansları ( $p < .01$ )

Madde Sayıları	Puanlayıcı Sayıları							
	2		3		4		5	
	Mutlak	Bağıl	Mutlak	Bağıl	Mutlak	Bağıl	Mutlak	Bağıl
9	.95870	1.02380	.81883	.80949	.54343	.57598	.46038	.48642
11	.78439	.78391	<b>.55788</b>	<b>.60128</b>	.44462	.47718	.37667	.40271
13	.66371	.72882	.50217	.54346	.37622	.40877	.31872	.34476
15	.57522	.64032	.43280	.46196	.32606	.35861	.27623	.30227
18	.47935	.54445	.34093	.38433	.27171	.30427	.23150	.26324

Tablo 11 incelendiğinde, matematik performansının ölçülmesinde kullanılan 11 maddelik klasik tarzdaki ölçme aracı ve üç puanlayıcıya göre elde edilen mutlak hata varyansı .55788 ve bağıl hata varyansı .60128 olarak kestirilmiştir. Madde sayısının iki

azaltılarak dokuz madde üzerinden; iki puanlayıcı için kestirilen mutlak hata varyansı .95870 ve bağıl hata varyansı 1.02380, puanlayıcı sayısının üç olması durumunda mutlak hata varyans değeri .81883 ve bağıl hata varyansı .80949, puanlayıcı sayısının bir arttırılıp dört olması durumunda .54343 ve bağıl hata varyansı .57598, puanlayıcı sayısının iki arttırılıp beş olması durumunda mutlak hata varyansı .46038 ve bağıl hata varyansı .48642 olarak kestirilmiştir.

Madde sayısının aynı kalması ve puanlayıcı sayılarının değişimlerine göre hata varyansları incelendiğinde; puanlayıcı sayısının iki azaltılarak iki olması durumunda mutlak hata varyansı .78439 ve bağıl hata varyansı .78391, puanlayıcı sayısının bir arttırılarak dört olması durumunda mutlak hata varyansı .44462 ve bağıl hata varyansı .47718, puanlayıcı sayısının iki arttırılarak beş olması durumunda mutlak hata varyansı .37667 ve bağıl hata varyansı .40271 olarak kestirilmiştir.

Madde sayısının arttırılarak 13 olması durumunda puanlayıcı sayısının azaltılarak iki olması durumunda mutlak hata varyansı .66371 ve bağıl hata varyansı .72882, puanlayıcı sayısının aynı kalması durumunda mutlak hata varyansı .50217 ve bağıl hata varyansı .54346, puanlayıcı sayısının bir arttırılması durumunda mutlak hata varyansı .37622 ve bağıl hata varyansı .40877, puanlayıcı sayısının iki arttırılması durumunda mutlak hata varyans değeri .31872 ve bağıl hata varyansı .34476 olarak kestirilmiştir.

Tablo 11' de yer alan değerler incelendiğinde madde sayısının ve puanlayıcı sayılarının artması durumlarında mutlak ve bağıl hata varyans değerlerinin giderek azaldığı açıkça görülmektedir. Tüm madde ve puanlayıcı sayılarının değişimi senaryolarına göre kestirilen mutlak ve bağıl hata varyansı değerleri incelendiğinde bağıl hata varyans değeri mutlak hata varyans değerine göre daha fazladır.



## **BÖLÜM V**

### **TARTIŞMA**

Bu bölümde, araştırma sürecinde toplanan veriler üzerinde yapılan istatistiksel analizler sonucunda elde edilen bulgular diğer araştırma sonuçları ile karşılaştırılarak tartışılmış ve değerlendirmelerde bulunulmuştur.

Bu araştırmada, ortaokul öğrencilerinin performanslarının belirlenmesinde kullanılan farklı ölçme araçları üzerinden aldıkları puanlar incelenmiştir. Bu inceleme yapılırken G Kuramı kullanılarak analizler yapılmış ve sonuçların birbirleri ile değişimleri gözlenmiştir.

Araştırma bulgularına göre, bireylerin çoktan seçmeli testten aldıkları notlar ile yazılı sınavdan aldıkları notların dağılımlarının paralellik gösterdiği gözlenmiştir. Çoktan seçmeli testten alınan notların daha yüksek olduğu ortaya çıkan bulgular arasındadır. Ranj değerlerinin değişimine baktığımızda klasik sınav için her bir puanlayıcının vermiş olduğu puanlar ile çoktan seçmeli teste ait ranj değerinin birbirleri ile çok yakın olduğu görülmektedir.

Klasik sınav ve çoktan seçmeli test için gerek ortanca gerekse standart sapma değerlerinin ortalama ekseninde değişimleri için belirlenen başarı notlarının çoktan seçmeli sınav için dağılımları ile paralellik gösterdiği görülmüştür. Ancak bu araştırmada başarı notları açısından çoktan seçmeli sınavdan alınan notların daha yüksek olduğu ortaya çıkmıştır.

Çetin (2009) yapmış olduğu araştırmasında; performans görevi, klasik sınav ve çoktan seçmeli sınav arasındaki ilişkiyi farklı değişkenlerle incelemiştir. Çetin, araştırma sonuçlarına göre başarı notlarının çoktan seçmeli sınav için daha yüksek olduğu sonucuna ve üç sınav arasında ilişkinin orta düzeyde olduğu sonucuna ulaşmıştır. Ancak ikili ilişkilere bakıldığında çoktan seçmeli sınav ile yazılı sınav arasındaki ilişkinin daha ileri düzeyde olduğu gözlenmiştir. Diğer ikili karşılaştırmalara göre, yapılmış olan bu araştırmada çoktan seçmeli sınav ve klasik sınav arasında ilişki yüksek bulunmuş; uygulama amacına göre sınavların uygulanmasında araştırmacının istediği özelliklere göre her iki sınavında kullanılabilirliği sonucuna varılmıştır. Klasik sınavda soru sayısının az olması gibi dezavantajlarının yanında puanlayıcılar arası

tutarlılığın sağlanması halinde çoktan seçmeli sınava yakın sonuçlar verdiği ortaya çıkmıştır.

Eser' in 2011 yılında yapmış olduğu sınav türlerini tercih etmeye ilişkin çalışmış olduğu betimsel tarama modelindeki araştırmasında öğrencilerin, başarı notları daha yüksek olduğu için çoktan seçmeli sınavları, yazılı sınavlara göre daha çok tercih ettiklerini belirtmişlerdir. Araştırma sonuçlarına göre en az tercih edilen sınav türü klasik sınav türü olarak belirtilmiştir. Yapılan bu çalışmada ise tercih türleri araştırılmamış ancak çoktan seçmeli sınav puanlarının dağılımlarının klasik sınav türünden elde edilen puan dağılımlarına göre daha yüksek olduğu sonucu ortak bulunan sonuç olmuştur.

Özellikle öğrencilerin bir şeyler üretmesini isteyen klasik sınavların öğrencilerde kaygı ve korkuya neden olduğu ve bu nedenle öğrencilerin başarılarının düşük olduğu farklı araştırmalarda ortaya konulmuştur. Ömür (2002) çalışmasında, öğrencilerin cevap üretmek yerine verilen cevaplar arasından birini seçmeyi daha çok tercih ettiklerini belirtmiştir. Ayrıca başarının yazılı sınavlarda çoktan seçmeli testlere göre daha yüksek olduğu sonucu bu çalışmada ortaya çıkan bir diğer bulgudur.

Bunun aksine bazı çalışmalarda klasik sınavda ortaya konulan performansın çoktan seçmeli sınavlara göre daha yüksek olduğu sonucuna ulaşan çalışmalar da vardır. Önder (2008), matematik başarısının ölçülmesi ve sınav kaygı düzeyi üzerine yapmış olduğu çalışmada; klasik sınava hazırlanan öğrencilerin başarılarının daha yüksek olduğunu belirtmiştir. Ayrıca çalışmada, hangi tür sorularla sınavlara hazırlanırsa hazırlansınlar, öğrencilerin klasik sınavlarda daha başarılı oldukları sonucu elde edilmiş; klasik sorularla sınava hazırlanan öğrencilerin yanı sıra, çoktan seçmeli sorularla sınava hazırlanan öğrencilerin de klasik sınavlarda daha iyi bir performans gösterdiği bulunmuştur. Oysa, bu araştırmanın bulgularından birisi öğrencilerin performans notlarının, çoktan seçmeli sınav için klasik sınava göre daha yüksek olduğudur.

Alan yazın incelendiğinde farklı sınav türlerinin karşılaştırıldığı ve üzerinde G Kuramı çalışması yapılan araştırmalara rastlanmamıştır. Daha çok performansın

belirlenmesinde puanlayıcıların birbirleri ile tutarlılığının incelendiği ve farklı desenlere göre karşılaştırılmaların yapıldığı araştırmalar mevcuttur.

Yapılan analizler sonucu ortaya çıkan bulgulara göre puanlayıcılar arasında KTK' ya göre yüksek korelasyonun ve G Kuramına göre puanlayıcı etkisinin düşük olması puanlayıcılar arasında yüksek uyum olduğu sonucunu ortaya çıkarmıştır.

Güler (2008) farklı kuramlara göre karşılaştırma yaptığı çalışmada; matematik başarısını belirlemede uygulanan klasik sınav verileri üzerinden KTK, G Kuramı ve ÇDRM çalışmaları yapmıştır. Elde edilen bulgulara göre G Kuramı çalışması sonuçlarına göre puanlayıcılar arasında tutarlılığı yüksek bulunmuştur. Nalbantoğlu, 2009 yılındaki araştırmasında; puanlayıcıların birlikte ve dönüşümlü olarak puanlamalarında sonuçlar arasında paralellik olduğu ve puanlamaların birbirleri ile tutarlı olduğu sonucuna ulaşmıştır.

LLabre 1978'deki çalışmada farklı modlar ve farklı yazma becerilerini değerlendiren puanlayıcıların vermiş oldukları puanların aradaki zaman ve farklı ortamlara rağmen tutarlı sonuçlar verdiği sonucuna ulaşmıştır. Puanlayıcı sayısının artması halinde güvenilirlik değerinin yükseldiği sonucu araştırmadan çıkan sonuçlardandır.

Çoktan seçmeli ve klasik sınavlarda madde sayısının artması sonucu güvenilirlik değerinin arttığı bulgularda gözlenmiştir. Puanlayıcı ve madde sayısının artması araştırmanın güvenilirliği açısından önemli bir özelliktir. Ancak uygulama, maliyet ve zaman gibi etkenlerden dolayı araştırmalarda hangisinin tercih edilebileceği hakkında bir noktaya varılmak istendiğinde bulgular dahilinde çoktan seçmeli sınav için madde sayısının artırılmasının; klasik sınav için puanlayıcının sayısının artırılmasının güvenilirlik değerlerini daha çok yükselttiği görülmektedir.

Her iki sınav türü içinde güvenilirlik çalışması yapılmış ve güvenilirlik indeksleri olarak KTK için  $\alpha$  ve G Kuramı için G katsayısı hesaplanmıştır. Araştırma için hesaplanan bu değerlere göre  $\alpha$  ve G katsayıları oldukça yüksek ve birbirlerine yakın bulunmuştur. Wang (2005) yaptığı araştırmasında benzer bir çalışma yapmış ve farklı

güvenirlik indekslerini hesaplamış ve karşılaştırmıştır. Çalışmanın sonucunda  $\alpha$  ve G katsayısının birbirine yakın olduğu sonucuna ulaşılmıştır.

## **BÖLÜM VI**

### **SONUÇ VE ÖNERİLER**

Bu bölümde, araştırmanın bulguları üzerinden elde edilen sonuçlar özetlenmiş ve bu sonuçlara dayalı yapılabilecek önerilere yer verilmiştir.

#### **6. 1. Sonuçlar**

##### **6. 1. 1. Çoktan Seçmeli Test İçin G Kuramına Göre Kestirilen Parametrelerin Varyansları Toplamı ve Taoplam Varyansları Açıklama Yüzdeleri Alt Problemine Ait Sonuçlar**

Birey (b) ana etkisinin tek değişkenli modelle kestirilen varyans değeri toplam varyans içinde ikinci yüksek değere sahiptir. Bireylerin maddelere vermiş oldukları cevapların farklılaştığı buradan da bireylerin verdikleri cevapların birbirlerinden kısmen bağımsız olduğu sonucuna varılmıştır.

Madde (m) ana etkisinin tek değişkenli modelle kestirilen varyans değeri toplam varyans değeri içinde en düşük değere sahip olduğu analiz sonuçlarına göre elde edilmiştir. Bulunan bu değere göre, uygulanan çoktan seçmeli sınava ait maddelerin güçlük düzeylerinin birbirlerine yakın olduğu dolayısıyla testin homojen bir yapı sergilediği sonucuna varılmıştır. Ayrıca bireylerin öğrenme düzeyleri birbirlerine yakın olduğu için maddelerin büyük bir boyutta değişmeye neden olmadığı sonucuna varılabilir.

Birey x madde ortak etkisinin toplam varyans içinde açıkladığı oran en yüksek değere sahip olduğu sonucuna ulaşılmıştır. İdeal bir test sınavında olması gerektiği gibi maddeler homojen, bireyler nispeten heterojen yapı sergilemektedir.

##### **6. 1. 2. Çoktan Seçmeli Test İçin Yapılan K Çalışması Doğrultusunda Maddelerin Hata Varyanslarının Değişimleri Alt Problemine Ait Sonuçlar**

Çoktan seçmeli test için yapılan K çalışmasına göre hesaplanan bağıl değerlendirme ve mutlak değerlendirme süreçleri için hata varyans değerleri sonuçlarına göre esas uygulama için mutlak hata varyans değerinin, bağıl hata varyans değerine göre daha yüksek olduğu sonucuna varılmıştır. Ayrıca çoktan seçmeli test için madde

sayısının arttırılıp azaltılması durumunda yapılan senaryolara göre, gerek bağıl gerek mutlak hata varyans değerlerinin azalma gösterdiği ancak mutlak hata varyans değerinin bağıl hata varyans değerinden daha yüksek olduğu ortaya çıkmıştır.

Bu sonuçlara göre, çoktan seçmeli test için yapılacak olan mutlak değerlendirmeye karışabilecek hata oranının bağıl değerlendirmeye karışabilecek hata oranından daha fazla olabileceği ve uygulanan sınavda madde sayısının artması durumunda değerlendirmeye karışan hata oranının azalabileceği sonucuna ulaşılmıştır. Bu bilgilerden yola çıkılarak, çoktan seçmeli testler için bağıl değerlendirme yapılmasının daha uygun olduğu ve çoktan seçmeli testlerde hata oranının daha az etkili olduğu sonucuna ulaşılmıştır.

### **6. 1. 3. Çoktan Seçmeli Test için Yapılan K Çalışması Sonuçlarına Göre G ve Phi Katsayılarının Değişimleri Alt Problemine Ait Sonuçlar**

Uygulanan çoktan seçmeli sınav için 22 madde üzerinden dönüştürülen puanlara ait genellenebilirlik (G katsayısı) ve güvenilirlik ( $\Phi$  katsayısı) birbirlerine çok yakın ve oldukça yüksek değerler olarak elde edilmiştir. Yapılan K çalışması sonucunda elde edilen değerlere göre, hem genellenebilirlik (G) katsayısı hem de güvenilirlik ( $\Phi$ ) katsayısı için madde sayısının arttırılması hem genellenebilirlik (G) hem de güvenilirlik ( $\Phi$ ) katsayılarının artmasını sağlayacağı sonucuna varılmıştır. Madde sayılarında küçük bir artma veya azalma durumlarında genellenebilirlik (G) ve güvenilirlik ( $\Phi$ ) katsayılarında az da olsa değişim meydana gelebileceği sonucuna ulaşılmıştır.

### **6. 2. 1. Klasik Sınav İçin G Kuramına Göre Kestirilen Parametrelerin Varyansları ve Toplam Varyansı Açıklama Yüzdeleri Yüzdeleri Alt Problemine Ait Sonuçlar**

Birey x madde x puanlayıcı (bmp) ortak yani artık etkisinin tek değişkenli modelle yapılan G Kuramı analizlerine göre toplam varyansın açıklama yüzdesi içerisinde en büyük etkiye sahip olduğu sonucuna ulaşılmıştır. Bulunan bu sonuç; varyans değerine göre birey, madde ve puanlayıcı ortak etkisinin büyük boyutta olduğunu ya da ölçmeye karışan tesadüfi hata oranının yüksek olduğunu göstermektedir.

Birey (b) etkisinin tek değişkenli modelle yapılan analizine göre en yüksek ikinci varyans değerini oluşturduğu sonucu ortaya çıkmıştır. Bunun bu varyans değeri araştırmalarda istenen bir durumdur. Bu sonuç; performansın ölçülmesinde birey

etkisinin varyans yüzdesi arasında büyük bir orana sahip olması bireyler arasında farklılıkların ortaya konabileceği şeklinde yorumlanır.

Birey x madde (bm) ortak etkisi tek değişkenli modelle yapılan analizine göre toplam varyans içinde en yüksek üçüncü değere sahip olmuştur. Birey x madde ortak etkisinin ,toplam varyans değeri içinde yüksek bir etkiye sahip çıkması bireylerin maddelere verdikleri cevapların birbirlerinden farklılık gösterdiği şeklinde yorumlanabilir.

Puanlayıcı (p) ana etkisi, tek değişkenli modelle yapılan analize göre toplam varyans içerisinde dördüncü sırada ve düşük bir etkiye sahip olmuştur. Buna göre, puanlayıcıların vermiş oldukları puan değerlerinin birbirlerinden farklılık göstermediği ancak bu etki çok yüksek değerde olmadığı için puanlayıcıların vermiş oldukları puanların birbirleriyle tutarlı olduğu sonucuna varılmıştır. Puanlayıcılar arasında korelasyonun çok yüksek olmasına rağmen hata varyansının büyük olmasının nedeni, öğrencilerin seviyelerinin birbirlerine yakın olması olabilir. Ayrıca puanlara verilen analitik rubrik, puanlayıcılar arasında farklılaşmayı engellemiş ve bu da toplam varyans içindeki varyansı açıklamayüzdesinin düşük çıkmasına neden olmuş olabilir.

Birey x puanlayıcı ( bp), madde (m) ve madde x puanlayıcı (mp) etkilerinin tek değişkenli modelle yapılan G Kuramı çalışmasına göre varyans değerlerinin sıfır olması bu araştırma için bu değişkenlerin ve etkileşimlerinin önemli bir etkiye sahip olmadığı sonucunu ortaya koymuştur. Varyans değerlerinin sıfır olması, yapılan araştırma için bu etki kaynaklarının değişimlerinin araştırma sonuçları etkileyen varyans kaynakları arasında olmadığı anlamını ortaya çıkarmıştır.

## **6. 2. 2. Klasik Sınav İçin Yapılan K Çalışmasına Göre Farklı Senaryolara Göre G ve Phi Katsayılarının Değişimi Alt Problemine Ait Sonuçlar**

Performansın ölçülmesine yönelik yapılan 11 madde ve bu maddeleri puanlayan üç puanlayıcıya göre kestirilen genellenebilirlik (G) ve güvenilirlik ( $\Phi$ ) katsayılarının birbirlerine yakın ve oldukça yüksek değerde olduğu sonucu elde edilmiştir. Yapılan K çalışması sonuçlarına göre, madde ve puanlayıcı sayılarının arttırılması durumunda hem genellenebilirlik (G) hem de güvenilirlik ( $\Phi$ ) katsayılarında artış meydana geldiği sonucuna varılmıştır. Madde sayısının aynı kalması durumunda puanlayıcı sayısında

artış meydana geldiğinde genellenebilirlik (G) ve güvenilirlik ( $\Phi$ ) katsayılarının değişiminin; madde sayısının arttırılması ve puanlayıcı sayısının aynı kalması durumu ile birbirlerine çok yakın olarak değişim gösterdiği sonucuna ulaşılmıştır. Ancak puanlayıcı varyansının toplam varyansa oranının madde varyansına göre daha yüksek olması puanlayıcı sayısının değişiminin genellenebilirlik (G) ve güvenilirlik ( $\Phi$ ) katsayılarını daha yüksek etkilediği sonucunu ortaya çıkarmıştır. Bunun içinde maliyetin ve araştırma süreci göz önüne alındığında ölçme aracında bulunun madde sayısının arttırılması ya da puanlayıcı sayısının arttırılmasına araştırmacının uygulama sürecine göre karar vermesinin yakın sonuçlar verebileceği sonucuna ulaşılmıştır. Araştırma sonuçlarına göre; puanlayıcı sayısı arttıkça güvenilirlik sayısı ve G katsayısının yüksek değerlere ulaşmak istenen araştırmalar için puanlayıcı sayısının arttırılmasının daha avantajlı olduğu sonucuna ulaşılmıştır.

### **6. 2. 3. Klasik Sınav İçin Yapılan K Çalışması Doğrultusunda Maddelerin Hata Varyanslarının Değişimi Alt Problemine Ait Sonuçlar**

Performansın ölçülmesinde 11 madde ve üç puanlayıcı üzerinden yapılan G Kuramı analizlerine göre elde edilen mutlak ve bağıl hata varyans değerlerinin birbirlerine yakın olduğu sonucu ortaya çıkmıştır. Puanlayıcı sayılarının ve madde sayılarının artması durumlarında hem mutlak hem de bağıl hata varyans değerlerinin azaldığı sonucuna ulaşılmıştır. Hata varyans değerlerine bakıldığında, bağıl hata varyanslarının her zaman mutlak hata varyans değerlerinden yüksek olduğu gözlenmiştir. Puanlayıcı sayısının artması durumunda hata varyans değerlerinin ( hem bağıl hem de mutlak hata varyansı için) madde sayısının artmasına oranla daha çok azaldığı görülmüştür.

Araştırmada tek değişkenli modelle yapılan G Kuramı analizlerinde kestirilen varyans yüzdesine göre; puanlayıcıya ait varyans yüzdesinin toplam varyansa oranının, madde değişkenine ait varyans yüzdesinin toplam varyansa oranının yüksek olması sonucu ile ilişkili olduğu sonucuna varılmıştır. Hata varyansının azaltılması istenildiğinde madde sayısının arttırılması veya puanlayıcı varyansının azaltılması için puanlayıcı sayısının arttırılması durumlarından puanlayıcı sayısının arttırılmasının daha fazla etkili olduğu sypılan araştırma sonucu ortaya çıkmıştır.



## 6. 2. Öneriler

Öneriler arařtırmadan ıkarılan ve yapılabilecek bařka arařtırmalar iin öneriler olmak üzere iki bařlık altında ařađıda verilmiřtir.

### 6. 2. 1. Arařtırmadan ıkan Öneriler

1. Performansın ölçülmesinde puanlayıcıların sayılarının farklı olması durumunda güvenirlilik deđerleri farklılık gösterebilir.
2. Mutlak veya bađıl deđerlendirme durumları dođrultusunda belirlenen yöntemlere göre karşılařtırma yapılabilir.
3. Öğrencilerin performansının karşılařtırılmasında oklu deđerlendirme yapılarak karşılařtırma yapılabilir.

### 6. 2. 2. İleride Yapılacak Arařtırmalar İin Öneriler

1. Performansın ölçülmesi iin farklı ölçme araçları kullanılarak kestirilen parametreler karşılařtırılabilir.
2. Yapılan G Kuramı analizlerinde farklı desenler kullanılarak farklı arařtırmalar yapılabilir.
3. alıřmanın yapıldığı konu alanı deđiřtirilerek farklı alıřmalar yapıp sonuçlar karşılařtırılabilir.

## KAYNAKÇA

- Açıkgöz, K. Ü. (2004). *Aktif öğrenme*. İzmir: Eğitim Dünyası Yayınları.
- Aiken, L. R. (1991). *Psychological testing and assessment* (7. baskı). Boston: Allyn and Bacon.
- Akdeniz, M. (2003). *Kısa cevaplı ve çoktan seçmeli testlerin madde ve test istatistiklerinin karşılaştırılması*. Yayınlanmamış yüksek lisans tezi. Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Alharby, E. R. (2006). *A comparison between two scoring methods, holistic vs. Analytic using two measurement models, the generalizability theory and the many facet rasch measurement within the context of performance assessment*. Unpublished doctoral dissertation. The Pennsylvania State University Faculty of Education, Pennsylvania.
- Alkahtani, S. F. (2012). *Oral performace scoring using generalizability theory and many-facet rasch measurement: a comparison study*. Unpublished doctoral dissertation. The Pennsylvania State University, Pennsylvania.
- Atılğan, H. (2004). *Genellenebilirlik kuramı ve çok değişkenlik kaynaklı rasch modelinin karşılaştırılmasına ilişkin bir araştırma*. Yayınlanmamış doktora tezi. Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Atılğan, H., Kan, A. ve Doğan, N. (2009). *Eğitimde ölçme değerlendirme*. Ankara: Anı Yayıncılık.
- Au, F., Prahardhi, S., & Shiell, A. (2008). Reliability of two instruments for critical assessment of economic evaluations. *Value in Health, 11*, 435- 439.
- Bahar, M., Nartgün, Z., Durmuş, S. ve Bıçak, S. (2008). *Geleneksel- alternatif ölçme ve değerlendirme teknikleri*. Ankara: Pegem Yayıncılık.
- Başol, G. (2008). Bilimsel araştırma süreci ve yöntem. İçinde Kılıç, O. & Cinoğlu M. (Ed.), *Bilimsel araştırma yöntemleri*, Bölüm 5, İstanbul: Lisans Yayıncılık.
- Başol, G. (2014). *Eğitimde ölçme ve değerlendirme*, Ankara: Pegem Yayıncılık.
- Bottema-Beutel, K., Lloyd, B., Carter, E. W., Asmus, J. M. (2014). Generalizability and decision studies to inform observational and experimental research in classroom settings. *American Journal on Intellectual and Developmental Disabilitie, Vol. 119, No. 6*, 589–605.
- Brennan, R. L. (2001). *Generalizability theory*. New York: Springer-Verlag.
- Briesch, A. M., Swaminathan, H., Welsh, M., Chafouleas, S. M., (2014). Generalizability theory: a practical guide to study design, implementation, and interpretation. *Journal of School Psychology, 52*, 13–35.
- Brophy, J. E. (1987). *On motivating students, in: D. Berliner & B. Rosenshine (Eds) talks to teachers*. New York: Random House.
- Brown, J. D., (2005). Generalizability and decision studies. *SHIKEN: JALT Testing & Evaluation SIG Newsletter. 9* (1), 12 – 16.

- Bruckner, C. T., Yoder, P. J., & McWilliam, R. A. (2006). Generalizability and decision studies: an example using conversational language samples. *Journal of Early Intervention*, 28, 139-153.
- Burns, K. J. (1998). Beyond classical reliability: using generalizability Theory to assess dependability. *Research in Nursing and Health*, 21, 83-90.
- Burton, E. B. (1998). *An investigation of the school-level generalizability of performance assessment results*. Unpublished doctoral dissertation. Rutgers University, New Jersey.
- Çakıcı Eser, D. (2011). *Genellenebilirlik kuramı ve lojistik regresyona dayalı hesaplanan puanlayıcılar arası tutarlığın karşılaştırılması*. Yayınlanmamış yüksek lisans tezi. Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Çelik, D. (2000). *Okullarda ölçme ve değerlendirme nasıl olmalı?*. İstanbul: MEB Yayınları.
- Çepni, S. (2009). *Araştırma ve proje çalışmalarına giriş*. Ankara: Pegem Yayıncılık.
- Çepni, S. ve Çil, E. (2009). *Fen ve teknoloji programı (Tanıma, Planlama, Uygulama ve Sbs'yle İlişkilendirme) ilköğretim 1. ve 2. kademe öğretmen el kitabı*. Ankara: Pegem Yayıncılık.
- Demir, E. (2010). *Uluslararası öğrenci değerlendirme programı (PISA) bilişsel alan testlerinde yer alan soru tiplerine göre Türkiye'de öğrenci başarıları*. Yayınlanmamış yüksek lisans tezi. Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Forsyth, D. R. (2003). *The professor's guide to teaching : Psychological principles and practices*. Washington, D.C. : American Psychological Association.
- Fry, H., Ketteridge, S., & Marshall, S. (2003). *A handbook for teaching & learning in higher education*. London: Kogan Page.
- Gao, X., & Brennan, R. L. (2001). Variability of estimated variance components and related statistics in a performance assessment. *Applied Measurement in Education*, 14, 191-203.
- Güler, N. (2008). *Klasik test kuramı, genellenebilirlik kuramı ve rasch modeli üzerine bir araştırma*. Yayınlanmamış doktora tezi. Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Güler, N., Kaya Uyanık, G. ve Taşdelen Teker, G. (2012). *Genellenebilirlik kuramı*. Ankara: Pegem Yayıncılık.
- Gültekin, M . (2005). *Öğretimde planlama ve değerlendirme*. Eskişehir: Anadolu Üniversitesi Yayınları.
- Haladyna, T. M. (1997). *Writing test items to evaluate higher order thinking*. Boston: Allyn and Bacon.
- Huang, J. (2012). Using generalizability theory to examine the accuracy and validity of large-scale ESL writing assessment. *Assessing Writing* 17, 123–139.
- Hocking, R. R. (1984), *Analysis of Linear Models*, Monterey, CA: Brooks-Cole.
- Huba, M. E., & Freed, J. E. (2000). *Learner-centered assessment on college campuses : Shifting the focus from teaching to learning*. Boston : Allyn and Bacon.

- Karasar, N. (1998). *Araştırmalarda rapor hazırlama yöntemi*. Ankara: Pars Matbaacılık.
- Karip, E. (Ed.) (2009). *Ölçme ve değerlendirme*. Ankara: Pegem Yayıncılık.
- Kaya, G. (2011). *Genellenebilirlik kuramının doldurma kavram haritası değerlendirme çalışmasına uygulanması*. Yayınlanmamış yüksek lisans tezi. Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Kumazawa, T. (2009). Revision of a criterion-referenced vocabulary test using generalizability theory. *JALT Journal*, 31(1).
- Lin, C. K., Zhang, J. (2014). Investigating correspondence between language proficiency standards and academic content standards: a generalizability theory study. *Language Testing*, Vol. 31(4), 413–431.
- MEB, 2007. *Merkezli eğitim uygulama modeli*. Ankara: Millî Eğitim Basımevi.
- Mert, V. (2008). *Enerji konusunda alternatif ölçme araçlarının geliştirilmesi*. Yayınlanmamış yüksek lisans tezi. Gazi Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Moon, S. Y. (1995). *Performance assessment: measurement issues of generalizability, dependability of scoring and relative information on student performance*. Unpublished doctoral dissertation. The Florida State University, Tallahassee.
- Mushquash, C., & O'Connor, B. P. (2006). SPSS and SAS programs for generalizability theory analysis. *Behavior Research Methods*. 38 (3), 542-547.
- Nalbantoğlu, F. (2009). *Performans ölçümlerinde genellenebilirlik kuramıyla farklı desenlerin karşılaştırılması*. Yayınlanmamış yüksek lisans tezi. Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill
- O' Neill ve O' Neill (2015), Improving QST reliability - more raters, tests, or occasions? A multivariate generalizability study, *The Journal of Pain*, Vol 16, No 5: pp 454-462.
- Özcan, F. (2011). *9. Sınıf öğrencilerinin kimyasal değişimler konusundaki bavrmsal Başarıları üzerine alternatif değerlendirme tekniklerinin etkisi*. Yayınlanmamış yüksek lisans tezi. Gazi Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Özçelik, A. (2011). *Fen ve teknoloji öğretmenlerinin ölçme- değerlendirme araçlarını kullanma sıklıkları ve karşılaştıkları sorunlar*. Yayınlanmamış yüksek lisans tezi. Dicle Üniversitesi Fen Bilimleri Enstitüsü, Diyarbakır.
- Özdemir, D. (1997). *Kısa cevaplı, çoktan seçmeli ve kompozisyon tipi maddelerden oluşan testlerin psikometrik özelliklerinin karşılaştırmalı olarak incelenmesi*. Yayınlanmamış yüksek lisans tezi. Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Öztürk, P. T. (2011). *İlköğretim 8. sınıf "canlılar ve enerji ilişkileri" ünitesinin kavram haritaları, yapılandırılmış grid ve tanılayıcı dallanmış ağaç teknikleri ile işlenmesinin öğrencilerin fen ve teknoloji dersine karşı tutumları üzerine etkisi*. Yayınlanmamış yüksek lisans tezi. Selçuk Üniversitesi Eğitim Bilimleri Enstitüsü, Konya.

- Polat, B. (2011). *Vee diyagramı, tanılayıcı dallanmış ağaç ve kavram haritalarının matematik dersine yönelik tutum ile başarıya etkileri ve bu araçlara ilişkin öğretmen görüşleri*. Yayınlanmamış yüksek lisans tezi. Hacettepe Üniversitesi Sosyal Bilimleri Enstitüsü, Ankara
- Rentz, J. O. (1987). Generalizability theory: a comprehensive method for assessing and improving the dependability of marketing measures. *Journal of Marketing Research*, 24(1), 19-28.
- Shavelson, R. J., & Webb, N. M. (1981). Generalizability theory: 1973–1980. *British Journal of Mathematical and Statistical Psychology*, 34, 133–166.
- Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. Newbury Park, CA: Sage.
- Shavelson, R. J., & Webb, N. M. (2005). Generalizability theory. Web: [http://web.stanford.edu/dept/SUSE/SEAL/Reports\\_Papers/methods\\_papers/G%20Theory%20A%20ERA.pdf](http://web.stanford.edu/dept/SUSE/SEAL/Reports_Papers/methods_papers/G%20Theory%20A%20ERA.pdf) adresinden alınmıştır.
- Semerci, Ç. (1992). *Fırat Üniversitesi' nde öğrenci başarısının ölçülmesinde kullanılan yöntemler ile ölçme- değerlendirme ilişkin görüşler*. Yayınlanmamış yüksek lisans tezi. Fırat Üniversitesi Sosyal Bilimleri Enstitüsü, Elazığ.
- Senemoğlu, N. (2002). *Gelişim öğrenme ve öğretim: kuramdan uygulamaya*. Ankara: Gazi Kitabevi.
- Solano-Flores, G., & Li, M. (2006). The use of generalizability (G) theory in the testing of linguistic minorities. *Educational Measurement: Issues and Practice*, 25(1), 13-22.
- Suen, H.K., & Lei, P.W. (2007). Classical versus Generalizability theory of measurement. *Educational Measurement*.
- Tekindal, S. (Ed.) (2011). *Eğitimde ölçme ve değerlendirme*, Ankara: PegemA Yayıncılık.
- Tosyalı, S. (2002). *Yetenek testlerinde çoktan seçmeli, kısa cevaplı ve doğru-yanlış testlerin madde ve test özelliklerinin karşılaştırılması*. Yayınlanmamış yüksek lisans tezi. Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Turgut, M. F. ve Baykul, Y. (2010). *Eğitimde ölçme değerlendirme*. Ankara: Pegem Yayıncılık.
- Wang, Z. (2005). *Estimating reliability under a generalizability theory model for writing scores in C-base*. Yayınlanmamış doktora tezi. University of Missouri, Columbia.
- Yalçın, M. (2012). *Lise kimya eğitiminde kullanılan farklı (yazılı) ölçme türlerinin çözelti konusunda öğrencilerin başarısını değerlendirmedeki güvenilirlikler*. Yayınlanmamış yüksek lisans tezi. Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Yelboğa, A. (2007). *Klasik test kuramı ve genellenabilirlik kuramına göre güvenilirliğin bir iş performansı ölçeği üzerinde incelenmesi*. Yayınlanmamış doktora tezi. Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.

## **EKLER**

**Ek 1: Ön Uygulama Yapılan Çoktan Seçmeli Test**

**Ek 2: Nihai Uygulama Yapılan Klasik Test**

### Ek 1: Ön Uygulama Yapılan Çoktan Seçmeli Test

- 1) 1.Kardeşin kan grubu O  
2.Kardeşin kan grubu BO  
3.Kardeşin kan grubu AO

Üç kardeşin kan grupları yukarıdaki gibidir. Buna göre anne ve babanın kan grubu genotipleri aşağıdakilerden hangisi olabilir?

	<u>Anne</u>	<u>Baba</u>
A)	B	O
B)	O	AB
C)	AO	BO
D)	AO	O

2) “AA,Aa ve aa” karakterleri,baskın ve çekinik olma durumuna göre hangisinde doğru eşleştirilmiştir?

	<u>Baskın Olanlar</u>	<u>Çekinik Olanlar</u>
A)	AA,Aa	aa
B)	AA	Aa,aa
C)	AA,aa	Aa
D)	Aa	AA,aa

3) Aşağıdaki karakterlerden hangileri çaprazlanırsa % 25 oranında buruşuk bezelyeler elde edilir? (D:düz bezelye, d:buruşuk bezelye)

- |            |            |
|------------|------------|
| A. DD X dd | B. Dd X Dd |
| C. DD x DD | D. DD x Dd |

4) Siyah saçlı anne ile siyah saçlı babanın sarı saçlı bir çocuğu olmuştur.

**Çocuğun sarı saçlı olmasında aşağıdakilerden hangisinin etkili olduğu söylenebilir?**

- A) Siyah saç geninin çekinik olması
- B) Anne ve babanın saf döl olması
- C) Anne ve babada sarı saç geninin olması
- D) Annenin saf döl babanın melez döl olması

5. Mitoz bölünmeyle ilgili olarak aşağıdaki verilenlerden hangisi yanlıştır?

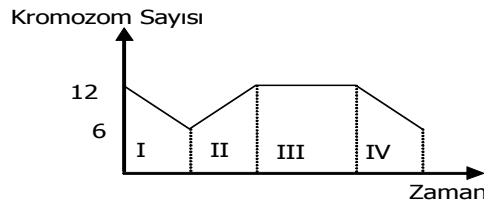
- A) Bir hücreli canlıların çoğalmasını sağlar
- B) Doku ve organlarımızdaki hücreler bölünür
- C) Kromozom sayısı değişmez
- D) Yeni hücrenin genetik yapısı farklı olur

- 6) I-Homolog kromozomlar yan yana gelir.  
 II- Parça değişimi görülür.  
 III- Kromozom sayısı değişmez.  
 IV- Üreme ana hücrelerinde görülür.

**Yukarıdaki ifadelerden hangisi mitoz bölünmeyi mayoz bölünmeden ayıran özelliklerden biridir?**

- A) I      B) II      C) III      D) IV

7) **Kromozom sayılarındaki değişim grafikte verildiği gibi ise, bu aşamalarda gerçekleşen olaylar aşağıdakilerden hangisinde doğru verilmiştir?**



- | I           | II       | III      | IV    |
|-------------|----------|----------|-------|
| A) Mitoz    | Mayoz    | Döllenme | Mitoz |
| B) Mayoz    | Mitoz    | Döllenme | Mayoz |
| C) Mayoz    | Döllenme | Mitoz    | Mayoz |
| D) Döllenme | Mayoz    | Mitoz    | Mitoz |

8) Hücre bölünmesi ile ilgili olarak;

- I. Hayvan hücresinde sitoplazma ara lamel oluşturarak bölünür.  
 II. Bitki hücresinde sitoplazma boğumlanarak bölünür.

**Yukarıdaki yargılar için ne söylenebilir?**

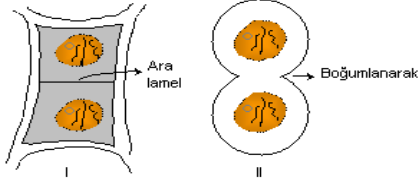
- A) I. yargı doğru, II. yargı yanlıştır.  
 B) I. yargı yanlış, II. yargı doğrudur.  
 C) I. ve II. yargı doğrudur.  
 D) I. ve II. yargı yanlıştır.

9)  **$2n=40$  kromozomlu bir ineğin bir hücresi ardı ardına iki mitoz ve bir mayoz bölünme geçirdiğine göre sonuçta oluşan hücre sayısı ve kromozom sayısı nedir?**

- A) 8 hücre, 20 kromozom  
 B) 16 hücre, 40 kromozom  
 C) 8 hücre, 40 kromozom  
 D) 16 hücre, 20 kromozom



10)



Kemal mitoz bölünme esnasında gerçekleşen sitoplazma bölünmesini poster olarak hazırlıyor. Bu postere bakarak; **aşağıdakilerden hangisi I ve II nolu bölünmeler için yanlış bir ifadedir?**

- A) I: Bitki hücrelerinde görülebilir.
- B) II: Hayvan hücrelerinde görülebilir.
- C) Oluşan hücrelerin kalıtsal özellikleri birbirinden farklıdır.
- D) Bölünme sonucunda 2 yavru hücre oluşur

11)



**Yukarıda verilen deniz yıldızındaki yenilenme olayı ile ilgili olarak aşağıdakilerden hangisi söylenebilir?**

- A) Eşeyli üretilir.
- B) Mitoz bölünme ile gerçekleşir.
- C) Yavru canlıların genotipi ana canlıdan farklıdır.
- D) Yavru canlılar ana canlıdan daha gelişmiş yapıdadır.

12) **Aşağıda verilenlerden hangisi kalıtsal hastalıklar arasında yer almaz?**

- A. Hemofili
- B. Lösemi
- C. Orak hücreli anemi
- D. Down sendromu

13) **Eşeysiz üremeye ilgili aşağıda verilen bilgilerden hangisi yanlıştır?**

- A. Anne ve baba hücreleri bir araya gelerek oluşur
- B. Tek hücreli canlılarda mitoz bölünme ile gerçekleşir
- C. Oluşan yavrular ana canlı ile aynı özelliktedir
- D. Denizanası, sünger eşeysiz çoğalan canlılardır

14) **Aşağıdakilerden hangisinde mutasyon yoktur?**

- A) Down sendromu
- B) Van kedisinin gözleri
- C) Altıparmaklılık
- D) Yaraların iyileşmesi

<b>Buğday</b>	<b>2n=42 kromozom</b>
<b>Soğan</b>	<b>2n=16 kromozom</b>
<b>İnsan</b>	<b>2n=46 kromozom</b>
<b>Eğrelti otu</b>	<b>2n=500 kromozom</b>
<b>Köpek</b>	<b>2n=78 kromozom</b>

15) Yukarıdaki tabloda bazı canlıların kromozom sayıları verilmiştir buna göre aşağıdakilerden hangisi doğrudur?

- A) Bitkilerin kromozom sayısı hayvanlarınkinden fazladır
- B) Kromozom sayısı ile canlının büyüklüğü orantılıdır
- C) Kromozom sayısı ile canlıların gelişmişliği arasında bir bağlantı yoktur
- D) Kromozom sayısı fazla olan canlı daha gelişmiştir

16) — Yumurta hücresinin spermle döllenmesi

- Hidranın tomurcuklanması
- Zigotun embriyoyu oluşturması
- Kertenkelenin kopan kuyruğunu yenilemesi
- Hamurun mayalanması
- Zedelenen derinin zamanla iyileşmesi

Yukarıdaki olayların kaç tanesinde mitoz bölünme vardır?

- A) 3
- B) 4
- C) 5
- D) 6

17) Mitoz bölünme ile ilgili aşağıdakilerden hangisi yanlıştır?

- A) Vücut hücrelerinde gerçekleşir.
- B) Yavru hücreler birbirinin tamamen aynısıdır.
- C) Yavru hücrelerin kromozom sayısı yarıya iner
- D) 2n kromozomlu ana hücreden 2n kromozomlu iki yavru hücre oluşur.

18) Bir köyde kalıtsal hastalıklı kişilerin oranında fazlalık olmasının nedeni aşağıdakilerden hangisi olabilir?

- A) İçme suyunun kirli olması
- B) Temizliğe dikkat edilmemesi
- C) Akriba evliliği oranının yüksek olması
- D) Yanlış veya fazla ilaç kullanımının yaygın olması

19) Bir insanda mayoz bölünme sonucunda oluşan aşağıdaki üreme hücrelerinden hangisinde mutasyon sonucu anormallik olmamıştır?

- A) 23+X
- B) 23+Y
- C) 22+XX
- D) 22+Y

20)	<u>Canlı Türü</u>	<u>Kromozom Sayısı</u>
	<b>Kedi</b>	<b>98</b>
	<b>Eğrelti Otu</b>	<b>2000</b>
	<b>İnsan</b>	<b>46</b>
	<b>Somon Balığı</b>	<b>46</b>

**Yukarıdaki tabloya göre;**

I.Kromozom sayıları ile canlıların büyüklüğü ve gelişmişliği arasında bir ilişki yoktur.

II. Kromozom sayısı fazla olan canlılar daha gelişmiştir.

III. Kromozom sayısı türden türe farklılık gösterebilir.

IV. Farklı türden canlılar aynı kromozom sayısına sahip olabilir.

**İfadelerinden hangileri doğrudur?**

A)I ve IV    B)II ve III    C)I,II ve III    D) I-III ve IV

**21) Aşağıdakilerden hangisi genetik mühendisliği çalışmalarının doğuracağı olumsuz sonuçlardan biridir.**

- A) Daha verimli tohumların elde edilmesi
- B) Daha kaliteli tarım ürünleri elde edilmesi
- C) Kişiye özgü ilaçların yapılması
- D) Farklı genetik hastalıkların ortaya çıkması

**22) 1800 Nükleotitten oluşan bir DNA zincirinde 400 tane Guanin olduğuna göre fosfat ve Timin sayısını bulunuz?**

- A) 900 fosfat, 500 Timin
- B) 1800 fosfat, 900 Timin
- C) 900 fosfat, 400 Timin
- D) 1800 fosfat, 500 Timin

**23) Bir DNA zincirinde toplam 2000 nükleotid vardır. Timin sayısı 700 ise Guanin sayısı kaçtır?**

- A)200    B)300    C)700    D)1000

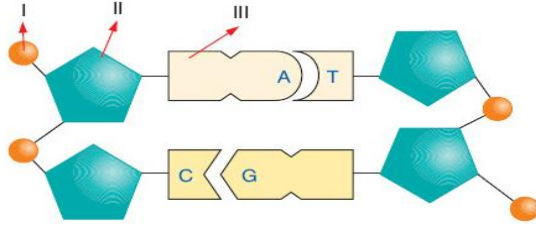
**24) Aşağıda verilen özelliklerden hangisi genlerle taşınarak nesiller boyu aktarılır?**

- A. Kesik parmaklı olmak    B. Güçlü kaslı olmak
- C. Kulak memesi bitişik olmak    D. Kısa saçlı olmak

**25) ” Kullanılan organlar gelişir kullanılmayanlar körelir.” Görüşü hangi bilim adamına aittir?**

- A) Wallace    B) Lamarck    C) Darwin    D) Pascal

26)



Yukarıda gösterilen bir nükleotid için ilgili kısımlar hangi seçenekte doğru olarak verilmiştir?

I	II	III
A) Baz	Şeker	Fosfat
B) Şeker	Fosfat	Baz
C) Fosfat	Şeker	Baz
D) Şeker	Baz	Fosfat

27) I. Gen

II. DNA

III: Nükleotid

Yukarıda verilen genetik yapının elemanlarının küçükten büyüğe doğru sıralanışı aşağıdakilerden hangisidir?

- A) I, II, III  
B) III, I, II  
C) II, I, III  
D) I, III, III

28) Aşağıdakilerden hangisi biyoteknoloji çalışmalarının hayatımızda sağlayacağı olumlu yönlere sahiptir?

- A) Çevre kirliliği sorunlarının incelenmesi  
B) Genetik çalışmalarla ürünlerin kalitesinin artırılması  
C) Biyolojik silah yapımında kolaylık sağlaması  
D) Makinelerin biyolojik faaliyetli yapılar haline getirilmesi

29) Aşağıdakilerden hangisi modifikasyonun tanımıdır?

- A) Canlının DNA'sında meydana gelen ani değişikliklerdir  
B) Canlının belirli bir yerde yaşama ve üreme şansını artıran kalıtsal özellikleridir  
C) Çevreye uyum sağlayan canlının hayatta kalması, uyum sağlayamayan canlının yok olmasıdır  
D) Çevrenin etkisiyle canlının dış görünüşünde meydana gelen kalıtsal olmayan değişikliklerdir

30) "Bireylerin çevre şartlarının etkisiyle sonradan kazandığı özellikler kalıtsaldır" diyen ve vücudu kashı olan birinin doğacak çocuğunun da kashı olacağını anlamına gelen bu sözü söyleyen bilim adamı kimdir?

- A) Darwin  
B) Mendel  
C) Lamarck  
D) Newton

**31) Aşağıdakilerden hangisinde adaptasyon vardır?**

- A) Çölle yaşayan tilki, fare ve tavşanların kulakları ve kuyrukları uzun ve vücut yüzeyleri geniştir
- B) Van kedilerinin gözlerinin renkleri farklıdır
- C) Saç renkleri siyah olan anne babanın çocuklarının saçları da siyah renklidir
- D) Kertenkelenin kopan kuyruğunu yenile

**32) Anadolu panteri ve Hazar kaplanı Anadolu'da yaşamış; 1970'li yıllarda nesli tükenmiş iki canlıdır.**

**Bu canlıların nesillerinin tükenmesinin sebebi aşağıdakilerden hangisi olamaz?**

- A) Doğal yaşam alanlarının azalması
- B) Av kaynaklarının yok olması
- C) Değişen ortam koşullarına uyum sağlamaları
- D) Besin bulmak için insanların yaşadığı yerlere yönelmeleri

**33) Aşağıda verilen seçeneklerden hangisi adaptasyona örnek olarak verilebilir?**

- A) Yılanların yaşadıkları ortama uygun renkte olması
- B) Albino ten rengi
- C) Ten renginin yazları bronzlaşması
- D) Down Sendromu hastası olmak

**34) Bir hücreden iki hücre oluşması için aşağıdaki olaylardan hangisinin gerçekleşmesi gereklidir?**

- A) DNA' nın kopyalanması
- B) DNA' nın parçalara ayrılması
- C) Nükleotid diziliminin değişmesi
- D) Hücrenin parçalanması

**35) Belli bir olgunluğa giren bireylerin vücutlarında üreme özelliğine sahip hücreler üretilmeye başlanır. Bu bilgiye göre hücrelerin oluşması aşağıdakilerden hangisi ile açıklanabilir?**

- A) Mitoz bölünme
- B) Tomurcuklanma
- C) Mayoz bölünme
- D) Dölllenme

- 36) X: Canlının görüntüsünü belirleyen genetik yapıya denir.  
Y: Canlının fiziksel özelliklerine verilen addır.

**Yukarıdaki bilgiye göre X ve Y aşağıdakilerden hangisidir?**

<u>X</u>	<u>Y</u>
A) Fenotip	Fenotip
B) Genotip	Genotip
C) Fenotip	Genotip
D) Genotip	Fenotip

- 37) I. Parça değişimi  
II. Döllenme  
III: Mayoz bölünme

**Yukarıda verilenlerden hangileri eşeyli üremede görülür?**

- A) Yalnız II B) I ve III C) II ve III D) I, II ve III

- 38) -Çuha bitkisinin çiçeğinin sıcaklığa bağlı olarak renk değiştirmesi  
-Yazın güneşli günlerde tenimizin bronzlaşması  
-Spor yapan kişilerin kaslarının gelişmesi

**Yukarıda verilen olaylar aşağıdaki ifadelerden hangisi ile açıklanabilir?**

- A) Mutasyon  
B) Modifikasyon  
C) Adaptasyon  
D) Evrim

- 39) I. Sperm hücresi II.Kan hücresi  
III.Yumurta hücresi IV.Amip hücresi

**Yukarıda verilenlerden hangilerinin oluşmasında mayoz bölünme görülür?**

- A) I ve II B) II ve III C) II ve IV D) I ve III

- 40) I. üreme hücrelerinde görülür  
II. kromozom sayısı yarıya iner  
III. parça değişimi oluşur  
IV. canlılarda çeşitliliğe neden olur

**Mayoz bölünme ile ilgili yukarıdaki ifadelerden kaç tanesi doğrudur?**

- A) 1 B) 2 C) 3 D) 4

## Ek 2: Nihai Uygulama Yapılan Klasik Test

1- Hücre bölünmeleri yaşamsal açıdan önemli bir biyolojik süreçtir. Buna göre aşağıda verilen sorular kısaca cevaplayınız.

- a) Mitoz bölünme hangi hücrelerde görülür ve neden önemlidir?
- b) Mitoz bölünme sonucu oluşan hücrelerin özellikleri nelerdir?
- c) Mayoz bölüme hangi hücrelerde görülür ve neden önemlidir?
- d) Mayoz bölünme sonucu oluşan hücrelerin özellikleri nelerdir ve bunun nedeni nedir?

2-Homozigot (saf döl) koyu renk saçlı bir anne ile Heterozigot (melez gen) koyu renk saçlı bir babanın çocuklarının saç renkleri nasıl olabilir? Çaprazlama yaparak gösteriniz. **(Koyu renk açık renge baskındır. Koyu renk: K, açık renk: k)**

3- Toplam 1600 organik baz olan bir DNA zincirinde 300 adet Adenin bazı varsa diğer organik bazların sayısını hesaplayınız.

4- Siyah gözlü anne ile siyah gözlü babanın mavi gözlü çocuğu olmuştur. Bu durum nasıl açıklanabilir?

5- Akraba evliliği neden sakıncalıdır?

6- Bütün canlıların DNA sı çift sarmallıdır ve aynı maddelerden yapılmıştır ( fosfat, deoksiriboz şekeri ve organik baz). DNA da dört çeşit organik baz vardır. Milyonlarca canlı çeşitliliği **nasıl** açıklanabilir?

7- Genetik mühendisliği ve biyoteknoloji çalışmalarının insan yaşamı açısından önemini açıklayınız.

8- Aşağıda verilen kavramların anlamlarını ve kavramlar hakkında birer örnek veriniz.

Kavran Adı	Kavramın Açıklaması	Örnek
Mutasyon		
Modifikasyon		
Adaptasyon		

9- Hücre bölünme safhalarını basit bir şekilde aşağıya çiziniz.

10- Adaptasyon ve evrim arasındaki ilişkiyi kısaca açıklayıp evrim hakkındaki görüşleri belirtiniz.

11- Genetik hastalıkların nedenlerinden kısaca bahsedip sık karşılaşılan genetik hastalıklara örnekler veriniz.



## ÖZGEÇMİŞ

<b>Adı Soyadı</b>	Muammer YÜKSEL
<b>Kişisel Bilgiler</b>	Uyruğu: T.C. Doğum Tarihi ve Yeri: 10.07.1988 / Manisa
<b>İletişim Bilgileri</b>	Tel: 0 506 846 54 12 E-posta: muammer.8845@gmail.com
<b>Öğrenim Bilgileri</b>	Lise: 2002–2006 Soma Rıfat Dağdelen Anadolu Lisesi Lisans: 2006–2010 Eskişehir Osmangazi Üniversitesi Eğitim Fakültesi İlköğretim Matematik Öğretmenliği Yüksek Lisans: Gaziosmanpaşa Üniversitesi Eğitim Bilimleri Enstitüsü Eğitimde Ölçme ve Değerlendirme Yüksek Lisans Programı Yüksek Lisans: Celal Bayar Üniversitesi Sosyal Bilimleri Enstitüsü Yönetim ve Organizasyon Yüksek Lisans Programı
<b>İş Deneyimi</b>	Soma Özel Linyit Çözüm Dershanesi Matematik Öğretmeni (2010-2011) Kaynarlı Ortaokulu (Selim, Kars) (2011-2014) Ayşe Temizel Ortaokulu (Soma, Manisa) (2014- ...)

