

T.C.
HASAN KALYONCU ÜNİVERSİTESİ
LİSANSÜTÜ EĞİTİM ENSTİTÜSÜ
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME ANABİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME TEZLİ YÜKSEK LİSANS
PROGRAMI

ATA MODEL UÇAK PUANLAMA ANAHTARININ GÜVENİRLİĞİNİN
BELİRLENMESİNDE GENELLENEBİLİRLİK KURAMININ KULLANIMI

YÜKSEK LİSANS TEZİ

HAZIRLAYAN

HASAN ÖZER

GAZİANTEP-2021

T.C.
HASAN KALYONCU ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME ANABİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME TEZLİ YÜKSEK LİSANS
PROGRAMI

ATA MODEL UÇAK PUANLAMA ANAHTARININ GÜVENİRLİĞİNİN
BELİRLENMESİNDE GENELLENEBİLİRLİK KURAMININ KULLANIMI

YÜKSEK LİSANS TEZİ

HAZIRLAYAN

HASAN ÖZER

DANIŞMAN

Doç. Dr. Ufuk AKBAŞ

GAZİANTEP-2021

TEZ ONAY SAYFASI



TEZ ETİK VE BİLDİRİM SAYFASI

Yüksek lisans tezi olarak sunduğum “Ata Model Uçak Puanlama Anahtarının Güvenirliğinin Belirlenmesinde Genellenebilirlik Kuramının Kullanımı” başlıklı çalışmanın tarafımda, bilimsel ahlak ve geleneklere aykırı düşecek bir yardıma başvurmaksızın yazıldığını ve yararlandığım eserlerin kaynakçada gösterilenlerden oluştuğunu ve bunlara atıf yapılarak yararlanmış olduğumu belirtir ve onurumla doğrularım. 15/08/2021

HASAN ÖZER



ÖNSÖZ

Tez çalışmam esnasında değerli bilgi, birikim ve tecrübeleriyle hep yanımda olan destekleriyle bana yol gösterici olup çalışmalarında beni sürekli motive eden çok değerli danışmanım Doç. Dr. Ufuk AKBAŞ'a teşekkürlerimi sunuyorum. Ayrıca lisans eğitimim süresince derslerine katılıp, bilgi birikimime çok büyük katkı sağlayarak bu alanda birlikte çalışmaktan büyük mutluluk duyduğum hocalarım, Prof. Dr. Şener BÜYÜKÖZTÜRK, Dr. Öğr. Üyesi Ersoy KARABAY, Doç. Dr. Yeşim ÖZER ÖZKAN, Doç. Dr. Halil İbrahim SARI ve Hasan Kalyoncu Üniversitesi Eğitimde Ölçme ve Değerlendirme Ana Bilim Dalı Araştırma Görevlisi Merve YILDIRIM SEHERYELİ'ne teşekkür ederim.

Akademik hayatım boyunca hep yanımda olup beni destekleyen değerli annem Suphiye ÖZER ve bu süreçte beni destekleyerek sürekli motive eden ismini yazamadığım ancak manevi desteklerini hiçbir zaman esirgemeyen geniş ailemin her bir ferdine sonsuz teşekkürlerimi sunuyorum.

Gaziantep, 2021

Hasan
ÖZER



Rahmetli babama sevgi ve özlemle...

ÖZET

Bu araştırmanın temel amacı; Genellenebilirlik kuramına göre Ata model uçakların birden fazla puanlayıcı tarafından Ata model dereceli puanlama anahtarı ile puanlanmasıyla elde edilen puanların belirlenen desenler ile Genellenebilirlik çalışmaları sonuçlarını incelemektir.

Araştırmada veriler, Gaziantep ilinde farklı okullarda eğitim alan 10 yaş ve üzeri öğrencilerin 32 saatlik eğitimleri sonucunda hazırlanan 120 uçağın 8 madde ve 5 puanlayıcının puanlaması ile verilerin toplanmasıyla elde edilmiştir. Elde edilen veriler Genellenebilirlik Kuramında tümüyle çaprazlanmış desen kullanılarak G çalışması ve K çalışması yapılmıştır.

Elde edilen sonuçlara göre, G çalışması sonucunda genellenebilirlik ve güvenilirlik katsayısının kabul edilebilir düzeyde olmadığı görülmüştür. Karar çalışmasında ise yapılan senaryolar sonucunda madde sayısının artırılması G katsayısını önemli oranda artıracığı sonucuna varılmıştır.

Anahtar sözcükler: Model uçak, genellenebilirlik kuramı, G çalışması, K çalışması, Dereceli puanlama anahtarı

ABSTRACT

The main purpose of this research is; According to the generalizability theory, it is to examine the results of the Generalizability studies with the determined patterns of the scores obtained by scoring the Ata model airplanes with the Ata model rubric by more than one rater.

The data in the research were obtained by collecting the data by scoring 8 items and 5 raters of 120 aircraft, which were prepared as a result of 32 hours of training for students aged 10 and over who were educated in different schools in Gaziantep. G study and K study were done by using the completely crossed design in the Generalizability Theory of the obtained data.

According to the results obtained, it was seen that the generalizability and reliability coefficients were not at an acceptable level as a result of the G study. As a result of the scenarios made in the decision study, it was concluded that increasing the number of items would significantly increase the G coefficient.

Keywords: Model airplane, generalizability theory, G study, K study, rubric

İÇİNDEKİLER

ÖNSÖZ	i
ÖZET	iii
ABSTRACT	iv
İÇİNDEKİLER.....	v
TABLolar LİSTESİ	viii
KISALTMALAR.....	ix
BİRİNCİ BÖLÜM	
GİRİŞ.....	1
1.1. Problem Durumu.....	1
1.1.1. Problem Cümlesi.....	6
1.1.2. Alt Problemler	6
1.2. Araştırmanın Amacı	7
1.3. Araştırmanın Önemi	7
1.4. Varsayımlar	8
1.5. Sınırlılıklar	8
1.6. Tanımlar.....	8
İKİNCİ BÖLÜM	
KAVRAMSAL ÇERÇEVE VE İLGİLİ ÇALIŞMALAR	9
2.1. Ölçme.....	9
2.1.1. Ölçmede Hata	10
2.1.1.1. Hata Türleri	12
2.1.1.1.1. Sabit Hata	12
2.1.1.1.2. Sistematik Hata	13
2.1.1.1.3. Tesadüfi (seçkisiz- rastgele- random) Hata.....	13
2.2. Ölçme Kuramları	14
2.2.1. Klasik Test Kuramı (KTK).....	15

2.2.2. Madde Tepki Kuramı (MTK).....	20
2.2.3. Genellenebilirlik Kuramı	22
2.3. Performans ve Ürünlerin Ölçülmesi.....	30
2.3.1. Dereceli Puanlama Anahtarı (Rubrik)	32
2.3.1.1. <i>Analitik Dereceli Puanlama Anahtarı</i>	33
2.3.1.2. <i>Bütünsel Puanlama Anahtarı</i>	34
2.4. İlgili Çalışmalar	35
ÜÇÜNCÜ BÖLÜM	
YÖNTEM	41
3.1. Araştırma Modeli	41
3.2. Çalışma Grubu	41
3.3. Veri Toplama Aracı	42
3.3.1. Model Uçak Dereceli Puanlama Anahtarı Geliştirilme Süreci.....	42
3.3.1.1. <i>Ölçütlerin Belirlenmesi</i>	43
3.3.1.2. <i>Maddeler İçin Uzman Görüşüne Başvurulması</i>	44
3.3.1.3. <i>Deneme Formunun Hazırlanması</i>	44
3.3.1.4. <i>Deneme Formunun Uygulanması</i>	45
3.3.1.5. <i>Nihai Formların Oluşturulması Uygulamanın Yapılması</i>	45
3.4. Verilerin Analizi	47
DÖRDÜNCÜ BÖLÜM	
BULGULAR VE YORUM	49
4.1. Birinci Alt Probleme Ait Elde Edilen Bulgular ve Yorumlar	49
4.2. İkinci Alt Probleme Ait Elde Edilen Bulgular ve Yorumlar	51
4.3. Üçüncü Alt Probleme Ait Elde Edilen Bulgular ve Yorumlar	51
BEŞİNCİ BÖLÜM	
SONUÇ VE ÖNERİLER	54
5.1. Sonuçlar.....	54

5.2. Öneriler	56
KAYNAKLAR.....	58
EKLER.....	62
EK-1. Ata model uçak puanlama anahtarı	62



TABLULAR LİSTESİ

Tablo 1. Tek yüzeyli ölçmeler için varyans kaynakları.....	28
Tablo 2. İki yüzeyli ölçmeler için varyans kaynakları.....	29
Tablo 3. Puanlayıcılara ait bilgiler.....	42
Tablo 4. Ata Model Uçak Puanlama Anahtarı.....	46
Tablo 5. 5 Puanlayıcıdan Elde Edilen Puanlara Ait Betimsel İstatistikler	47
Tablo 6. $u \times m \times p$ Çaprazlanmış Desenine Ait G Çalışması Sonucunda Kestirilen Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri	49
Tablo 7. $u \times m \times p$ tümüyle çaprazlanmış desenine göre G ve Phi katsayıları	51
Tablo 8. $u \times m \times p$ Tümüyle Çaprazlanmış Desenine Göre Puanlayıcı veya Madde Sayılarının Artırılmasıyla Yapılan K Çalışması Sonucu Elde Edilen Bulgular	52

KISALTMALAR

α	: Cronbach Alfa Katsayısı
b	: birey
b x m	: birey-madde ortak etkisi
b x p	: birey-puanlayıcı ortak etkisi
b x p x m	: birey-puanlayıcı-madde ortak etkisi (artık)
G Çalışması	: Genellenebilirlik Çalışması
G Katsayısı	: Genellenebilirlik Katsayısı
G Kuramı	: Genellenebilirlik Kuramı
K Çalışması	: Karar Çalışması
KTK	: Klasik Test Kuramı
m	: madde
MKE	: Madde Karakteristik Eğrisi
MTK	: Madde Tepki Kuramı
Phi Katsayısı	: Güvenirlilik Katsayısı
p	: puanlayıcı
p x m	: puanlayıcı-madde ortak etkisi
SH	: Standart Hata

BİRİNCİ BÖLÜM

GİRİŞ

1.1. Problem Durumu

Türkiye’de son yıllarda ekonomik, siyasal ve kültürel alanlardaki dönüşümlerin bir kısmı eğitim alanında da gözlenmektedir. Millî Eğitim Bakanlığı (MEB) son yıllarda eğitim alanında bazı önemli değişiklikler yapmaya çalışmaktadır. 2005-2006 eğitim-öğretim yılından bu yana benimsenmiş bulunan ve öğretimden çok öğrenme odaklı bir özelliğe sahip olan “yapılandırmacı yaklaşım” modeline bağlı olarak özellikle ders programları ve içeriklerinde bazı önemli değişiklikler yapılmaya çalışılmıştır. Yapılandırmacı yaklaşıma dayanan yeni öğretim programları, derslerin içerik ve işleyiş biçimlerine ilişkin olarak hem kitapları hazırlayanlara hem de öğretmenlere daha esnek ve özgür olabilme fırsatları sunmaktadır. Yapılandırmacı eğitim yaklaşımıyla evrensel düzeydeki değişim ve gelişmeleri eğitim sistemine yansıtmayı amaçlayan yeni eğitim programlarının geliştirilmesi üzerine çalışılmaktadır. Talim Terbiye Kurulu Başkanlığı bu yaklaşımın benimsenmesindeki asıl amacının devamlı bir değişim içinde olan dünyayı takip edebilecek, üretilen bilgi ve birikime ulaşabilecek ve kullanabilecek, bireysel farklılıkları ne olursa olsun, araştırma-sorgulama, eleştirel düşünme, problem çözme ve karar verme becerileri gelişmiş vatandaşlar yetiştirmek olduğunu belirtmektedir (TTKB, 2005: 8-16).

Bir eğitim programının amacı, bireyin almış olduğu dersin öğretim programında belirlenen hedeflenen davranışların veya kazanımların kazandırılmasıdır. Hedeflenen kritik davranış veya kazanımların kazanılmış olma derecesinin belirlenmesi amacıyla, amacına göre farklı ölçme işlemleri yapılmaktadır. Buradan anlaşılacağı üzere ölçme herhangi bir durumda gerçekleşen bir durumu gözlemleyip gözlem sonucunun var olan olanaklarla sayısal bir biçimde gösterilmesidir(Özçelik, 2013: 1). Bununla birlikte ölçme işlemi farklı şekillerde tanımlanmıştır.

Crocker ve Algina (2006: 5) Psikolojik bir özelliği ölçmek için kullanılan ölçme aracıyla davranışa niceliksel bir değer atamasıyla ölçme gerçekleşir. Eğitimde ve psikolojide yapılan ölçümlerde temel amaç benzer davranışları belirlemek için davranışların ölçülerek onları tanımlamak ve sınıflamaktır. Daha açıkça ifade etmek gerekirse ölçme, psikolojik yapılar arasında gözlemlenebilir ilişkiler ortaya koymaya çalışır. Ölçme aracılığıyla

arařtırmaların doęrulanmasıyla davranıřların tahmin edilmesi veya kontrol edilmesi mümkündür.

Baykul (2015: 85-86) ise ölçmeyi “gerçek dünyada var olan varlıkların veya elemanların niteliklerinin sayı veya semboller yardımıyla ifade ettięi bir model” olarak ele almaktadır.

Eęitim sisteminin geliřmiř olduęu ölkelere bakıldıęında buralarda yařayan bireylerin yazılı yoklama veya çoktan seçmeli testler gibi ölçme araçlarına alıřık oldukları görölmektedir. Ölçme araçları, eęitimin her kademesinde kullanıldıęı gibi iře alım gibi durumlarda kullanılmakta ve önemli bir yer tutmaktadır. Öğrencilerin eęitim seviyesi ve beklenen performanslarını göstermelerinde bir uyarıcı görevi alan ölçme araçları, bireylerin başarı düzeyleri hakkında dönüt sağlar. Bu açıdan bakıldıęında ölçme araçları eęitimciler için önem arz etmektedir. Ölçme araçlarından elde edilen verilere dayanarak bireyler hakkında önemli kararlar alınabilmektedir. Ařaęıda ifade edilen durumların ölçme araçlarından elde edilen sonuçlarla iliřkili olduęu söylenebilir.

1. Gelecekle ilgili planlanan bir eęitim için, bireylerin bilgi ve becerileriyle ilgili kazanımlarının ölçülmesi.
 2. Bireylerin farklı programlar dâhilinde sınıflandırılması.
 3. Eęitim programı, iř ve ürün vb. durumlar için seçilecek adayların geleceęe yönelik performanslarının yordanması.
 4. Bireylerin, daha önceden belirlenmiř ve yetenek standartlarına sahip olup olmadıęına karar verme.
 5. Özel bir öğretim metodu veya eęitim programının etkili olup olmadıęını belirleme.
- řeklinde çoęaltılabilir(Tekindal, 2017: 2-3).

Son yıllarda özellikle yeni öğretim programlarında performans ölçme ve çeřitli öğretim ve öğrenme durumlarından söz edilmektedir. Önerilen tamamlayıcı ölçme araçları ve yöntemlerinin temel amacı eęitimde istenilen insan niteliklerini ölçerken daha önce kullanılan dięer ölçme araçlarıyla ölçülemeyen bazı niteliklerin ölçülmesinin giderilmesidir. Tamamlayıcı ölçme araç veya yöntemleri incelendięinde performans ölçümlerinin önemli bir yeri olduęu görölmektedir. Dilimize batı dillerinden giren “Performans” sözcüğünün sözlük anlamları řu řekildedir:

- Bir eylemin gerçekleştirilmesi durumu veya ortaya ürün konulması.

- Bir vaadin iddianın veya talebin yerine getirilme durumu.
- Sahne oyunlarında bir karakterin canlandırılması.

şeklinde ifade edilebilir(Özçelik, 2013: 74-76).

Yeni öğretim programlarında ele alınan performansların ölçülmesi, daha önce alışagelmış yöntem veya tekniklerle bireyin öğrenme düzeylerini belirlemek amacıyla yapılan sınavlara alternatif olarak gösterilmektedir. Bu yaklaşım, bireylerin performanslarının farklı yönlerini aynı anda ve birlikte ölçmesine imkân sağladığı izlenimi yaratmaktadır. Tamamlayıcı ölçme araçları ve yöntemlerinde; “*Projeler, Performans Ölçümleri, Seçilmiş Ürün Örnekleri*” başlıkları altında ele alınabilir(Özçelik, 2013: 76-77).

Belirlenen bir duruma göre performansın ölçülmesine karar verilmeden önce, performansı ölçmek için gerekli olan işlemler düşünülerek maliyetleri tanımlanmalıdır. Ekstra zaman ve gayret isteyen öğrenme sonuçlarıyla performansı ölçmek için ihtiyaç duyulan kaynaklar arasında istenileni seçmek ve etkili ölçmeler yapabilmek için ölçme yöntemleri bulmak ya da geliştirmek ve onların güvenilir ve geçerli olmasını sağlamak oldukça önemlidir. Performansı ölçme, doğrudan gözlem ölçüldüğünden kompleksli bilişsel, duyuşsal devinişsel ve sosyal becerilerin kullanımını gerektiren eğitimsel amaçlara hizmet etmesi gerekmektedir. Bir eğitimcinin öğretimini gerçekleştireceği ortamda performans ölçümlerini yapabilmesi için oluşturacağı ölçme araçlarının basamakları verilmiştir. Bunlar;

- Ölçülecek performansın tanımlanması.
- Derecelendirme kurallarını ve ölçütlerinin belirlenmesi.
- Performans çalışmalarını ve uygulamalarını otaya koymak.
- Gerekli ölçme aracını tasarlamaktır(Tekindal, 2017: 248).

Performansa dayalı durumların belirlenmesi yakın zamanda ortaya çıkan bir konu olduğu için, performans gerektiren zihinsel süreçlerin puanlanmasında, dereceli puanlama anahtarları (Rubrikler) kullanılmakta ve bu alanda araştırmalar devam etmektedir. Performans ölçümlerinde yapısal özellikleri bakımından iki farklı dereceli puanlama anahtarı kullanılmaktadır. Bunlardan ilki analitik dereceli puanlama anahtarı bir diğeri ise bütünsel (holistik) dereceli puanlama anahtarıdır(Kutlu, Doğan ve Karakaya, 2017: 53-54).

Öğrencilerin ortaya koyduğu performanslarının süreç ve sonuçlarını analiz etmek ve yorumlamak için öğretmenler veya ölçme uzmanları tarafından geliştirilen puanlama şemalarına dereceli puanlama anahtarı denmektedir. Puanlama öncesinde bu şekilde bir

şemanın ortaya konulması, yapılacak bir performans ölçümünün nesnel bir şekilde yapılmasını sağlar. Nesnel bir ölçüm için dereceli puanlama anahtarında hangi puanların hangi özelliği taşıyan yanıtlara karşılık geldiği açıkta tanımlanmalıdır. İki farklı dereceli puanlama anahtarlarından bütünsel dereceli puanlama anahtarı, bireylerin her bir göreve veya maddeye gösterdiği performansın bütünsel olarak, alt boyutlarına ayırmadan kullanılan ölçme aracıdır. Analitik dereceli puanlama anahtarı ise bütünsel dereceli puanlama anahtarının aksine bireyin her bir madde veya görevde ortaya koyduğu performansı öncelikle alt bileşenlerine ayırarak puanlamayı sağlar(Güler, 2019: 81-83).

Yapılan bütün işlemlerde ölçmenin bu ölçütlerin merkezinde olduğu ifade edilebilir. Bu sebeple belirlen ölçütlerin karşılaşmasında ölçmenin önemli bir yeri vardır. Ölçme işleminin fen bilimlerinde yapılması sosyal bilimlere göre daha kolaydır. Fen bilimlerindeki değişkenlerin genellikle kararlı özelliklere sahip olup fiziksel değişkenler olmasından dolayı doğrudan ölçümler yapılabilmektedir. Sosyal bilimlerde ise genellikle psikolojik değişkenlerle çalışılmakta ve dolaylı ölçümler yapılmaktadır. Bireylere bağlı psikolojik yapılarıdaki değişkenler doğrudan ölçülemeyi için, ölçme işlemlerinde birçok zorlukla karşılaşmaktadır. Bu sebeple bireylerin örtük özelliklerini en uygun şekilde ortaya çıkarabilmek çeşitli kuramlar geliştirilmiştir. İlk olarak ortaya konulan Klasik Test Kuramı (KTK) karşımıza çıkmaktadır. Daha sonra KTK'nın yetersizliğine karşı çıkan eleştiriler sonucunda Madde Tepki Kuramı (MTK), Çok Boyutlu Madde Tepki Kuramı (ÇBMTK) ve KTK'nın bir uzantısı kabul edilen Genellenabilirlik Kuramı (GK) ortaya çıkmıştır (Erkuş, Sünbül, Ömür Sünbül, Yormaz ve Aşiret, 2017: 107).

Bu ölçme kuramları hemen her yaş düzeyinde pek çok bilişsel ve duyuşsal özelliklerin ölçülmesinde, farklı puanlama türlerinde, gerek bilgisayar ortamında gerekse yüz yüze veya kâğıt-kalem testlerinde çeşitli amaçlarla kullanılmaktadır. Yapılan bu araştırmada da hazırlanan dereceli bir puanlama anahtarının kullanılmasıyla öğrenciler tarafından ortaya konulan performans sonucunda hazırlanan Ata model uçakların Genellenebilirlik Kuramına göre güvenilirlik kestirimleri yapılacaktır. Bu amaçla model uçak eğitimi verilmiştir ve bu eğitim sonucunda her bir öğrencinin Ata model uçağını tamamlanması sağlanmıştır.

Model uçak, havadan ağır motorlu veya motorsuz, insan taşıma kapasitesi olmayan, sınırlı boyutlarda imal edilmiş hava araçlarıdır. Model uçak rehber öğretmenleri tarafından 10 yaş ve üzeri öğrenenlere havacılığı sevdirmek ve özendirmek amacıyla belirli okullarda

eđitimciler tarafından birok eđitim verilmektedir. Eđitim erevesinde đrencilerin yapmıř oldukları model uakları hediye edilerek kurslar tamamlanmaktadır(THK, 2017).

Model uak eđitimlerinde đrencilerle eđitimler gerekleřtirilirken eđitmenlerin rehber rolünde olması nem arz etmektedir. Bu srete đrenciler model uak projelerindeki yapım ařamalarını adım adım takip ederek model uak rnlerini ortaya ıkarmaktadırlar(THK, 2017).

Milli Eđitim Bakanlıđı'nın (MEB) Hayat Boyu đrenme Genel Mdrlđ (HBGM) tarafından Halk Eđitim merkezleri (HEM), eđitmenleri tarafından 10 yař ve zerindeki bireylere Trkiye'nin birok ilinde 32 saatlik eđimler verilerek ocukların Ata model uak eđitimi almaları sađlanmaktadır. Bu eđitim programında kursiyere;

1- đrenci/kursiyerin yařam boyu kullanabileceđi ve mesleki geliřmesine yararlı olabilecek, retken, bilim ve teknoloji retimine yatkın, beceri dzeyi yksek olarak yetiřtirilmesi, iyi iliřkiler kurabilmesi, iře uyum sađlayabilmesi gibi genel bilgi ve becerilerin kazandırılması,

2- Bu ereveler ile đrenci/kursiyerlere ulusal ve uluslararası iř gcnden beklenen yeterlikleri kazandırılması,

3- THK' nın kurumsal yapısını ve tarihini, askeri ve sivil havacılıđın nemini kavrayarak dođru olarak uygulamalar yapabilmesi,

4- Kurallara uygun olarak serbest uuř ata model uađı yapabilmesi,

5- Temel kurallara uygun olarak model uak planını okuyabilmesi,

6- Uygun malzeme kullanarak lsne ve standartlara uygun olarak kanat yapabilmesi,

7- Plana uygun olarak ykseliř ve dnř dmeni yapabilmesi,

8- Plana uygun olarak model uađın gvde inřaatını yapabilmesi,

9- Uygun malzemeleri kullanarak model uađın kaplamasını yapabilmesi,

10- Model uak montajını yapabilmesi,

11- Model uçağın uçuş ayarlarını yaparak uçuşunu yapabilmesi, amaçlanmaktadır(HBÖGM, 2017: 3).

Ölçme ve değerlendirme aşamasında ise bireylerin, çeşitli ölçme araçları kullanılarak;

1. Kursun amaçlarına ulaşip ulaşmadığı ölçülecektir.
 2. Kursiyerlerin yaptığı tüm öğrenim faaliyetleri değerlendirilecektir.
 3. Değerlendirme; teorik ve/veya uygulamalı olarak yapılmalıdır.
 4. Teorik değerlendirme, programın tüm amaçlarının kazanılıp kazanılmadığını ölçecek şekilde yapılmalıdır.
 5. Değerlendirme 100 puan üzerinden yapılmalı ve sınav başarısı Yaygın Eğitim Kurumları Yönetmeliği esaslarına göre belirlenmelidir.
- şeklinde ifadelere yer verilmiştir(HBÖGM, 2017: 8).

1.1.1. Problem Cümlesi

Yukarıdaki maddelerde ifade edilmesine rağmen daha önce Ata model uçak modelini değerlendirmek için bir ölçme aracı ortaya konulmamıştır. Sonuç olarak, bakıldığında eğitim hayatında bireylerden istenen performansların ortaya konulması ve bu performansların ölçülerek değerlendirilmesi, bunun sonucunda da bireylerin ortaya koydukları performans veya ürüne göre güvenilir puanlar alması önem arz etmektedir. Bu nedenle araştırmada Ata model uçak puanlama anahtarının geliştirilip güvenilirliğinin test edilmesi amaçlanmıştır.

Bu amaçla araştırma temel olarak “Ata model dereceli puanlama anahtarından elde edilen puanların genellenebilirlik kuramına göre güvenilirliği nasıldır?” Sorusu çerçevesinde şekillenmiştir. Bu kapsamda öğrencilerin tamamlamış oldukları 32 saatlik eğitim sonucunda yapmış oldukları model uçak ürünlerinin 5 puanlayıcı tarafından puanlanması sonucunda puanlayıcılar arasında bir farklılık olup olmadığı sorusuna cevap aranmaktadır.

1.1.2. Alt Problemler

Genellenebilirlik Kuramı'nda uçak (u), madde (m) ve puanlayıcı (p) değişkenlerin tümüyle çapraz tasarlandığı (u x m x p) desende analitik dereceli puanlama anahtarı kullanarak elde edilen puanların;

1. G çalışması sonucunda kestirilen varyans bileşenleri ve toplam varyansı açıklama yüzdeleri nasıl değişmektedir?
2. G çalışması sonucunda kestirilen G ve Phi katsayı değerleri nedir?
3. Puanlayıcı sayılarının ve madde sayılarının değişmesi durumlarında; K çalışmasında kestirilen G katsayısı ile bağıl ve mutlak hata varyansları nasıl değişmektedir?

1.2. Araştırmanın Amacı

Bu araştırmanın genel amacı bireylerin performansına yönelik güvenilir puanlamalar yapılmasında puanlayıcılar arasındaki tutarlılığı inceleyerek puanlama esnasında puanlayıcı, uçak ve maddelerden kaynaklanan hata kaynaklarının belirlenip buna göre yapılacak bir puanlama için gerekli olan puanlayıcı sayısının belirlenmesini sağlamaktır. Araştırmada bireylerin ortaya koymuş olduğu performansla yönelik hak ettiği puanı alması ve puanlayıcılardan kaynaklanan hata oranının en aza indirilmesi önem arz etmektedir. Ayrıca bu araştırma sonucunda test geliştiricileri ve model uçak eğitmenleri için elde edilecek güvenilirliği ve geçerliği test edilmiş bir model uçak dereceli puanlama anahtarının kazandırılması hedeflenmektedir.

1.3. Araştırmanın Önemi

Model uçak eğitim uzmanlarının, bireylerin performansının veya ortaya koymuş oldukları ürünlerinin çok yönlü olarak değerlendirmesi beklenmektedir. Bu sebeple model uçak eğitmenlerinin bireylerin performansının değerlendirilmesinde adil puanlama yapabilmesi önem arz etmektedir. Yapılan bu çalışmada Model Uçak Dereceli Puanlama Anahtarı'nın 5 farklı puanlayıcı tarafından puanlanmasıyla elde edilen puanların güvenilirliklerinin Genellenebilirlik Kuramına göre incelenmesiyle elde edilen puanlama aracının bütün Ata model uçağı eğitimlerinde kullanılması adına doğru bir adım olacaktır.

Genel olarak yapılan ölçümlerin hatalardan arınık olması beklenmektedir. Bu çerçevede puanlama anahtarından hatalar ortaya çıkacağı gibi puanlayıcıdan da ortaya

çıkabilecek hata kaynakları ortaya çıkabilmektedir. Bu sebeple puanlayıcıdan kaynaklanan hataların da test edilmesi ve güvenilirliklerinin ortaya çıkarılması önem arz etmektedir.

1.4. Varsayımlar

- Model uçak eğitimi alan bireylerin el becerileri ve yaş düzeyleri model uçak yapmaya uygundur.

- Puanlama yapacak olan puanlayıcıların almış oldukları eğitim bütün puanlayıcılar tarafından aynı şekilde algılanmıştır.

1.5. Sınırlılıklar

- Model uçak eğitimi alıp model uçak ürününü ortaya koyan bireylerin yaşları 10 ile 14 yaş aralığındadır.

- Model uçak eğitimi alan bireylerin model uçaklarını tamamlamaları 32 saat süreli olarak belirlenmiştir. 32 saatlik eğitim süresinde uçaklarını tamamlayamayan bireylerin model uçakları değerlendirmeye alınmamıştır.

1.6. Tanımlar

Bu alanda araştırmada söz edilen bazı terimlerin tanımlamaları yapılmıştır.

Ölçme: Eğitimde ve psikolojide bireysel veya gruplar halinde insanlar hakkında karar almak veya insanların kendileri hakkında kararlar almaya yardımcı olması amacıyla kullanıldığını ifade etmektedir.

Puanlama anahtarı: Bir bireyin bir kavram, durum veya olayla ilgili bilgisini belirlemek veya bir çalışmaya ilişkin yeterliliğini belirlemek için kullanılan bir puanlama sistemi veya ölçme aracıdır. Buna ek olarak puanlama anahtarları, öğrencilerin zayıf ve güçlü yönleri ile ilgili geri bildirim verme, iyileştirme ve düzenleme amaçlı öğretimi planlama gibi nedenlerle de kullanılma alanları vardır.

İKİNCİ BÖLÜM

KAVRAMSAL ÇERÇEVE VE İLGİLİ ÇALIŞMALAR

2.1. Ölçme

Ölçmenin hem günlük hayatımızda, hem de bilim alanında yapılan çalışmalarda önemli bir yeri vardır. Bilimsel alandaki çalışma ve uygulamalarda, o bilim dalına ait ölçme araçlarının ve yöntemlerinin bulunmasıyla ölçme alanda çalışmalar hız kazanmıştır. Ölçme o kadar önemli bir hale gelmiştir ki, çeşitli bilim dalları için farklı ölçme teorileri geliştirilmiştir(Baykul 2015: 81).

Buradan yola çıkarak literatür incelendiğinde, ölçme ile ilgili birçok tanım yapıldığı görülmektedir.

Thorndike (2014: 3-10) ölçmenin eğitimde ve psikolojide bireysel veya gruplar halinde insanlar hakkında karar almak veya insanların kendileri hakkında kararlar almaya yardımcı olması amacıyla kullanıldığını ifade etmektedir. Buna göre ölçme teknikleri insanların daha iyi kararlar almasına yardımcı olabilir. Ölçmeler karar vermez. İnsanların karar vermesinde yardımcı olur. Yapılan ölçme işlemlerinde şu 3 adıma dikkat edilmelidir:

1. Ölçülecek nitelik veya özellik
2. Ölçülecek özellik veya niteliğin gözlemlenebilir olması
3. Gözlem veya kuralların niceliksel olarak ifade edilmesi.

Crocker and Algina (2006: 45) ölçmeyi nesnelerin özelliklerine belirli kurallar dâhilinde sayı veya semboller ataması olarak değerlendirmektedir. Deneysel gözlemleri ölçmek için sistematik kuralların ve anlamlı ölçü birimlerinin olması gerekmektedir.

Tekindal (2017: 20) “Daha önceden belirlenen standartların mevcut olaylar veya özelliklerle karşılaştırarak standartlaştırma işidir.” Şeklinde ifade etmiştir. Buradan yola çıkarak aynı özellikleri ölçen bireylerin benzer sonuçlar elde edilmesinin beklendiğine değinmektedir.

Özçelik (2013: 1) ise ölçmeyi herhangi bir durumda var olanı gözlemleyip, bu gözlem sonucunu sayısal olarak ifade etme işlemi şeklinde ele almıştır.

İstatistikte önemli bir kavram olarak görülen ölçme, gözlemlenen bir olayın sınıflandırılması veya bir değer verilmesi ile sonuçlanan işlemler şeklinde ele alınmaktadır(Köklü, Çokluk ve Büyüköztürk, 2016: 7).

Bu tanımlar incelendiğinde ölçme işleminin farklı şekillerde gerçekleştirilebileceği görülmektedir. Crocker and Algina'ya (2006: 14) göre, bireylerin davranışlarının için iyi tanımlanmış bir yapıya ve bu yapının ortaya konulması için bir ölçme süreci gerekmektedir. Yapılan ölçme çalışmalarında istenilen sonuçlar elde edilmesinde 5 temel sorunla karşılaşıldığını şu şekilde ifade etmektedir:

1. Bir özelliği ölçmek için evrensel olan tek bir yol yoktur.
2. Psikolojik ölçümler davranış örneklerini temel alır.
3. Yapılan tüm ölçme işlemlerinde hata yapılma ihtimali vardır.
4. İyi tanımlanmış bir ölçme birimi yoktur.
5. Psikolojik yapıların ölçümlerinin anlamlı olabilmesi için diğer değişkenlerle ilişkilerinin olması gerekmektedir.

Tespit edilen bu ölçme sorunlarına bakıldığında hata kaynaklarının ölçme işlemi yapan kişiden, ölçülen özellik veya yapıdan, kullanılan ölçme aracından ve ölçme işlemine kadar birçok çeşidinin olduğu görülmektedir. Ölçme işlemi yapılırken bu hata kaynakları ve ölçme işlemine karışan hatların bilinmesi, ölçmede daha kesin sonuçlara yaklaşan sonuçlar elde edilmesini sağlar. Ancak hiçbir zaman bir ölçmede kesin sonuçlar elde edildiği ifade edilememektedir(Güler, 2019: 22-23).

2.1.1. Ölçmede Hata

Bir testin güvenilirliği sorulduğunda, önemli olan testin neyi ölçtüğü değil, neyi doğru ölçüp ölçmediğidir. Ölçme sonucunda ortaya çıkan puanın kesinliği nedir? Tekrar bir ölçme işlemi gerçekleşirse aynı puan elde edilebilir mi? Evde yapılan bir ölçme işleminde alınan bir puan ile okulda aynı ölçme aracıyla yapılan bir ölçme arasında puanların aynı olmadığı ve puanlar arasında bir tutarsızlık olduğu görülmüştür. Bu ortaya çıkan tutarsızlık sonucunda kullanılan testte bir güvenilirlik sorununun olduğu görülmektedir. Yapılan ölçme işlemleri sabit olmasına rağmen test puanlarının farklı olmasından yola çıkarak Gerçek puana (T_i), o anda gözlenen puanlara veya ölçümlere (X_i) ve buna karışan hataların (E_i) karıştığı söylenebilir. Yani gerçek puan ile gözlenen puan arasındaki herhangi bir farka ölçme hatası denir. Birçok

ölçme işleminde ölçme hatası rastgele görülebilmektedir ve bu hatalar gözlenen puana pozitif yönde veya negatif yönde etki edebilir(Thorndike, 2014: 77).

Gerçekleştirilen bütün ölçme çalışmalarında hatanın olması kaçınılmazdır. Her ne kadar dikkatli ve doğru bir ölçümler yapılmaya çalışılsa da yine de ölçmede bir miktar hatanın bulunması söz konusudur. Hata genellikle şu şekilde ifade edilebilir:

$$\text{Hata} = \text{Gözlenen ölçme sonucu} - \text{Gerçek ölçme sonucu}$$

Bu denklemde görüldüğü gibi gözlemlenen ölçme sonuçlarının gerçek ölçme değerinden meydana gelen ölçme farkına hata denilmektedir. Her varlığın ölçülmek istenen özelliğine bağlı gerçek bir değeri vardır. Fakat bazen ölçme aracından, bazen ölçmeyi yapan bireylerden veya başka sebeplerden dolayı tam olarak doğru bir ölçüm yapılamaz. Ölçme işleminde ölçülecek nesne veya varlığın miktarının az veya çok olacak şekilde ölçülmesi şeklinde ölçümler yapılması hata yapılmasına sebep olmaktadır. Buradan anlaşılacağı üzere hatanın hem miktarı hem de bir yönü vardır, denilebilir (Güler, 2019: 22).

Ölçme işlemini yapan bireyin en önemli amacı hatasız bir ölçme elde etmektir. Ancak düşünüldüğü gibi hatasız bir ölçme elde etmek mümkün değildir. Doğrudan ölçme yapılan işlemlerde de hatalar yapılmaktadır(Tavşancıl, 2010: 13-14).

Fen bilimlerinde doğrudan gözlenen nesnelerin gözlemlenebilmesine rağmen, yapılan ölçme işlemlerinde az miktarda da olsa hatalar bulunmaktadır. Bu hatalar ölçme işlemini yapan kişi, ölçülen özellik veya ölçme aracından kaynaklanabilir. Dolaylı ölçmelerde doğrudan ölçmenin hatalarını barındırmak ile birlikte, ölçme işlemini gerçekleştiren kişinin fikir yapısı, bilgi birikimi, ölçme esnasındaki sağlık veya psikolojik durumu, gibi ek hata kaynakları içermektedir (Güler, 2019: 22; Karagöz, 2017: 22; Tavşancıl, 2010: 13-14).

Okullarda yapılan sınavlardan elde edilen puanlar, öğrencilerin başarı, bilgi gibi görülmeyen özelliklerinden elde edilen puanlardır ve öğrencilerin bu özelliklerinden elde edilen puanların gerçek puan olması mümkün değildir. Öğretmenlerinde puanlamış oldukları bu sınav sonuçlarına hatalar karışmış olabilir. Bu sınav sonuçları öğrencilerin gerçek puanını temsil etmez. Gerçek puanla birlikte bir miktar hata barındırır(Güler, 2019: 22; Tekindal,2017: 47).

Ölçme işlemlerinde ölçülmek istenen özelliğin tam değeri bilinmediğinden diğer bir deyişle her ölçme işlemine bir miktar hata karıştığından dolayı bu hata payının tam miktarını

belirlemek güçtür. Ancak ölçülen özelliğin kararlı olması diğer bir ifadeyle tekrarlı ölçümler sonucu benzer sonuçlar vermesi, hata miktarının ne kadar olduğu üzerinde ortalama bir tahminde bulunulabilir(Tavşancıl, 2010: 13-14).

Ölçülmek istenen hata değerinin sıfır olması beklense de pratikte bu mümkün değildir. Çünkü ölçme işleminde ölçmeye etki eden tüm faktörlerin kontrol altına alınması pratikte mümkün görünmemektedir(Karagöz, 2017: 22-23).

Bütün bunlar incelendiğinde hata kaynaklarının bazılarının şunlar olduğu söylenebilir. Bunlar:

- Ölçme işlemini gerçekleştiren puanlayıcının dikkatsizliği, psikolojik veya sağlık durumu, açlığı ve susuzluğu, puanlamada yanlı davranması gibi etkenler örnek gösterilebilir.
- Ölçme işleminin yapıldığı bireylerin sınav kaygısı veya korkusu, rahatsızlığı, şans başarısı ve kopya çekmesi gibi durumlar olabilir.
- Ölçme işleminin yapıldığı ortamın gürültülü olması, ortamın istenilenden soğuk veya sıcak olması, aydınlatmanın yetersiz olması gibi durumlar olabilir.
- Kullanılacak ölçme aracının açık uçlu testler olması sebebiyle puanlayıcının subjektif olması, şans başarısının yüksek olduğu testlerin kullanılması, testin kapsam geçerliliğinin düşük olması gibi birçok etkenden dolayı hatalar ortaya çıkabilir(Arıkan vd., 2014: 88-89; Can, 2018: 47; Erkuş vd., 2017: 9-10; Güler, 2019: 23; Karagöz, 2017: 23-24).

2.1.1.1. Hata Türleri

Daha önce ifade edildiği gibi ölçme işlemini yapan kişiden, ölçme aracından veya ölçülen özellikten olduğu gibi bunların dışında da birçok hata ölçmeye karışabilir. Bu açıdan literatür incelendiğinde hata türlerinin 3 farklı şekilde gruplandığı görülmektedir.

2.1.1.1.1. Sabit Hata

Her bir ölçme işlemi yapıldığında; ölçme sonucuna aynı yönde ve aynı miktarda yansıyan, diğer bir ifadeyle ölçme işlemlerinde bütün ölçme sonuçlarına aynı miktarda karışan hatalardır. Bu tür ölçme işlemlerinde hata kaynağı bellidir.

Fen bilimleri gibi fiziksel bilimlerden örnek verilecek olursak; 20 cm'lik bir cetvelin 1 cm'lik ucu kırılmış olsun. Bu cetvelle yapılan bir ölçümde ölçme işlemine hata karışır. Hatanın miktarı ve yönü aynıdır ve kaynağı da cetveldir.

Eğitim bilimlerinde yapılan sınavlarda sınav puanlarına puanlayıcı tarafından 10 puan eksik veya fazla verilmesi yine sabit hataya örnektir. Burada hatanın yönü ve miktarı aynıdır, hatanın kaynağı ise puanlayıcıdır.

2.1.1.1.2. *Sistemik Hata*

Sistemik hatalar, tekrarlı ölçme sonucunda miktarı ve yönü değişen ancak hata kaynakları belli olan hatalardır. Ucu kırık olan cetvelin birçok kez kullanılmasıyla birden fazla sonucun elde edilmesi buna örnektir. Aynı şekilde puanlayıcının yapılan bir yazılı sınavda, yazısı güzel olan veya daha çok yazı yazana ek olarak 5 puan vermesi gibi. Eğitimde puanlayıcının yapmış olduğu bu durumdaki hatalara yanlılıkta denmektedir.

Sistemik hatalarda hatanın miktarı ölçme aracının özelliğine, testi alan bireylerin ölçülen özellikleri dışındaki bazı özelliklerine veya ölçülen özelliğin miktarına göre değişebilmektedir. Buradan yola çıkarak sistemik hata kaynaklarının puanlayıcı veya ölçme aracı olduğu söylenebilir.

Fiziksel bilimlerde sistemik hatalarda hatanın kaynağının puanlayıcı veya ölçme aracından kaynaklandığı bilindiği için hatanın düzeltilmesi mümkün olabilmektedir. Ancak eğitim bilimlerinde yapılan ölçümlerde bu tür hataları belirlemek veya düzeltmek mümkün olmayabilir.

2.1.1.1.3. *Tesadüfi (seçkisiz- rastgele- random) Hata*

Eğitim bilimlerinde en çok üzerinde durulan hatalardır. Tesadüfi hatalar, yönü ve miktarı belli olmamakla birlikte hata kaynaklarının da belli olmadığı hatalardır. Bu tür hataların pozitif veya negatif yönde miktarı bilinmediği ve kaynağı belli olmadığından dolayı kontrol edilmesi imkânsızdır. Sabit ve sistemik hatalardaki gibi matematiksel işlemlerle düzeltilmesi mümkün değildir.

Tesadüfi hatalar puanlayıcının dikkatsizliği, yorgun olması gibi elinde olmayan durumlardan dolayı yapılan bir sınavdaki sınav sonuçlarını yanlış toplaması olabilir. Aynı şekilde öğrencinin katıldığı sınavlarda konuyu çok iyi bilmesine karşın rahatsızlanması veya başka sebeplerden dolayı alması gereken puandan farklı puanlar alması tesadüfi hatalara örnek verilebilir.

Sabit ve sistematik hatalar geçerliği, tesadüfi hatalarda güvenilirliği etkilediği için eğitim bilimlerinde en çok üzerinde durulan hatalar tesadüfi hatalardır. Bu sebeple ölçme de hata kuramları en çok tesadüfi hatalar üzerinde kurulmuştur (Arıkan vd., 2014: 90-91; Baykul, 2015: 100-101; Büyüköztürk vd., 2018: 108; Güler, 2019: 23).

2.2. Ölçme Kuramları

Ölçme, bireylerin özelliklerini temsil etmenin bir yolu olarak sistematik bir şekilde kişilere sayısal niceliklerin atanmasıdır. Bu nicelikler, dikkatle belirlenmiş, tekrarlanabilir bir işleme göre atanmalıdır. Belirli bir test, her konu için aynı talimatları, soruları ve puanlama işlemlerini kullanarak puanlar üretmektedir. İnşa edilecek ölçek türüne bağlı olarak çeşitli formlar kullanılarak ölçülecek özelliklere göre sayılar atanabilir. Yani, sayılar deneklere nominal ölçekler oluşturan kategorilerde, sıralı ölçekler oluşturan sıralarda veya aralık veya oran ölçeklerini oluşturan sürekli puanlar olarak atanabilir.

Ölçme kuramları, ölçümlerin kalitesini tanımlamaya, sınıflandırmaya ve değerlendirmeye çalışır. Ölçme araçlarının kullanılabilirliğini, doğruluğunu ve anlamlılığını geliştirmeye yönelik çalışmalar yapar. Yeni ve daha iyi ölçme araçları geliştirmek için yöntemler önerir. Sosyal ve davranış bilimlerindeki çeşitli disiplinler, kendilerine eşlik eden bir etiketle kendi benzersiz ölçme yaklaşımlarını benimsemiş veya yaratmıştır. Günümüz ölçme kuramlarının temellerinin çoğu 1950'lerde tamamlanmış olsa da, psikolojik ölçme (veya psikometriye) yönelik daha yeni yaklaşımlar için araştırma ve geliştirmeler devam etmektedir. Belirtildiği gibi, psikometri, işletme ve ekonomide kullanılan ekonometrinin aksine, davranış bilimleri ve eğitimdeki ölçme yaklaşımlarına hâkim olma eğilimindedir(Lester, Inman and Bishop, 2014: 45).

Bu açıklamalardan yola çıkarak bireylerin sahip oldukları bu örtük özelliklerinin uygun bir şekilde ortaya konulabilmesi için birçok kuram geliştirilmiştir. Klasik test kuramı

(KTK) 'nın ilk olarak geliştirilmesi ile birlikte süreç içerisinde bu kuramın eksikliklerine yönelik çeşitli eleştiriler yapılmıştır. Bu eleştiriler sonucunda Madde Tepki Kuramı (MTK) ortaya konulmuştur. Daha sonra ise Genellenabilirlik Kuramı (G kuramı) da aynı şekilde KTK'ın eksikliklerine bağlı olarak ortaya çıkmıştır(Erkuş vd., 2017: 107).

2.2.1. Klasik Test Kuramı (KTK)

Bireylerin bir özelliğini ölçmeye çalışırken ölçülen özelliklerine ölçme, esnasında etki eden bazı durumların etki ettiği ifade edilmektedir. Örneğin bir testte bireyin bir özelliği ölçülmeye çalışılırken, bireyin çabukluk ve uyanıklık durumu onun bu özellik hakkında gerçek bilgisinin ölçülmesine engel olabilmektedir. Bu sebeple böyle durumların göz ardı edilmesi kişinin ölçülen bilgisinin veya zekâsının eksik yorumlanmasına dolayısıyla ölçme işlemine hatanın karışması sebep olabilmektedir. Bunun dışında yorgunluk, duygu ve dikkatsizlik gibi durumlarda bireyin yeteneğinin gerçek değerinin bulunmasına engel teşkil etmektedir. Bu ve buna benzer durumlar kontrol edildiğinde ancak istenilen bir ölçme işlemi yapılabilmektedir(Spearman, 1927: 246).

Ölçme yoluyla gözlenen bir özelliğin gerçek değeri bulunmak istenir. Fakat ölçme işlemi esnasında ölçmeye karışan çeşitli hatalar gerçek değerini doğrudan ölçülmesine engel olmaktadır. Bu sebeple doğrudan ölçülemeyen gerçek değer, gözlenen puanlar yardımıyla kestirilmeye çalışılır. Doğru kestirimler yapabilmek için istatistiksel yöntemler aracılığıyla bazı kuramlar geliştirilmiştir. Bu kuramlardan biri de Klasik Test Kuramı (KTK)'dır. (KTK), eğitim ve psikoloji alanında gerçek dünyada karşılaştığımız yapıları, ölçme yoluyla kuramsal yapılara bağlayarak açıklamamızı sağlamaktadır. Bu sebeple KTK, bir ölçme kuramı olarak ele alınmaktadır. KTK'da ölçülmek istenen değişkenin gerçek değeri gerçek puan olarak isimlendirilmektedir. Gerçek puan doğrudan elde edilemediği için bazı varsayımlar, gözlenen puanlar yardımıyla kestirilmeye çalışılır. Bu nedenle KTK, Gerçek Puan Kuramı olarak da ifade edilebilmektedir(Baykul, 2015: 91).

Psikolojik ölçmelerde ölçme hatası fiziksel ölçmelerdekinden çok daha büyük olabilir. Bir testteki hata tahminini ve hata puanı ile gerçek puan arasındaki farkı bulmak psikoloji ve eğitimdeki ölçmelerin temel problemi olarak ele alınmaktadır. Gerçek puan ile hata puanları arasındaki ilişkiyi belirlemek için Gulliksen (1950: 4-5) bazı varsayımlar ileri sürmektedir. Buna göre uygulanan bir testten yola çıkarak:

Xi: kişinin testten gözlenen puanı

Ti : kişinin gerçek puanı

Ei: aynı kişinin hata puanı

olmak üzere: $Xi = Ti + Ei$ veya $Ei = Xi - Ti$ denklemi ortaya çıkmaktadır. Bu denklem gerçek puan ile hata puanı arasındaki ilişkiyi belirten varsayım olarak ele alınabilmektedir. Diğer bir ifadeyle gözlenen puandan farklı olarak ele alınan gerçek puan kabul edildiğinde bu iki puan arasındaki farkın hata puanı olarak adlandırıldığı ifade edilebilmektedir.

Lord and Novick (1968: 28-29) de hatasız bir ölçme işleminin gerçekleşmesini şu şekilde ifade etmektedir: Xi bir testten alınan puan olmak koşuluyla Ei kadar ölçme hatası içeriyorsa Xi-Ei 'nin hatasız bir ölçüm olduğunu ifade etmektedir. Bu da gerçek ölçüm olarak ele alınabilir. Gerçek ölçüm, gerçek değer olarak da adlandırılır. Buradan yola çıkarak gerçek değer doğrudan ölçülemeyeceği sonucuna varılabilir.

KTK genel olarak hata üzerine kurulan bir kuramdır. Dolayısıyla güvenilirlikle de doğrudan ilişkilidir. Magnusson'un (1967) güvenilirlik beklentisi tekrarlı ölçümler sonucunda benzer sonuçlar elde edilmesi üzerinedir. İfade edilen benzeşmeyi incelemenin teorik açıdan iki yolu bulunmaktadır. Bunlardan ilki, aynı kişiler üzerinde aynı ölçme araçlarıyla birden fazla ölçme işlemi yaparak tekrarlı ölçümler yapmaktır. Ancak psikoloji ve eğitimde ölçme konusu insan özellikleri olduğundan dolayı bir bireye bir test aynı anda birden fazla kez uygulanamamaktadır. Bu testlerin uygulanma durumu söz konusu olsa bile testteki aynı soruların sorulması, testi alan bireyin daha önce testteki soruları hatırlaması sebebiyle yapılan ölçümler birbirinden bağımsız olmamaktadır. Bu sebeple tekrarlı ölçme işlemleri yerine, paralel ölçmelerden yararlanır. Paralel ölçümler ile bir birey üzerinde tekrarlı ölçümlerin model veya uygulama üzerinde getireceği olumsuz sonuçlar yerine, bir grup üzerinde yapılan ölçümler ile kişilerin niteliklerinin kestirilmesi sağlanabilmektedir. Paralel ölçümler KTK'nın pratikte uygulanmasını sağlasa da kurama ait bazı çıkarımların yapılabilmesi için bazı varsayımların kabul edilmesi gerekmektedir. Bu varsayımlar şunlardır:

1. Bir evrenden yapılan rastgele ölçümlerin hata puanlarının beklenen değeri, başka bir ifadeyle ortalaması sıfırdır.

2. Bir evrenden yapılan rastgele ölçümler için hata puanları ile gerçek puanları arasındaki korelasyon sıfırdır.

3. Bir evrenden yapılan paralel iki ölçüme ait hata puanları arasındaki korelasyon sıfırdır.

4. Bir evrenden yapılan iki ölçüme birine ait olan hata puanı ile diğerine ait olan gerçek puan arasındaki korelasyon sıfırdır(Erkuş vd., 2017: 109-110; Baykul, 2015: 93-97; Crocker and Algina, 2006: 111; Lord and Novick, 1968: 36-37).

Varsayımlar incelendiğinde ölçme işleminde ortaya çıkan hata puanları ile gerçek puan arasındaki bağlantı KTK'nın temelini oluşturmaktadır. KTK da yukarıda ifade edilen varsayımların dışında ölçülmek istenen değişkenlerin tek boyutlu olması ve bireylere uygulanan testlerin paralel olması da varsayımlardan bazılarıdır(Erkuş vd., 2017: 109-111).

Paralel ölçümler için iki ölçmenin hem gerçek puanlarının eşit olması hem de hata puanları varyanslarının eşit olması gerekmektedir. Diğer bir ifadeyle paralel ölçmelerin aynı değişkenleri ölçmesi koşuluyla aritmetik ortalamaları ve varyansları eşit olan ölçmelerdir. Aynı şekilde iki testin paralel test olabilmesi için de test ortalamalarının ve standart sapmalarının eşit olması gerekmektedir. bu şekilde paralel testler arasındaki korelasyon güvenilirlik indeksini verecektir.(Baykul, 2015: 113-114).

Crocker and Algina (2006: 131-134) pratikte bir test geliştiricisinin pratikte mükemmel paralel ölçümler yapamayacağını ve bireylerin paralel ölçümler ile gerçek puanlarının elde edilemeyeceğini ifade etmektedir. Aynı şekilde bir dizi ölçümde bu puanlar için güvenilirlik katsayısının tam olarak belirlenemeyeceğini, belirli bir testte yanıt veren bireyler için tahmin edilebilmektedir. Testin güvenilirliğini tahmin etmek ve paralel ölçümlerin elde edilebilmesi için şu öneride bulunmuşlardır.

1. Paralel formlar yöntemi: Hazırlanan bir testin iki benzer formunun oluşturulması ve her iki formunun da sınava giren aynı bireylere uygulanmasını gerektiren yöntemdir. İki gruba ayrılan katılımcılardan ilk gruba birinci form ve diğer gruba ikinci form verilerek daha sonra arada yeterli zaman bırakılıp sınava girenlerin yorulmamasını sağlayarak formların değiştirilmesi ile her iki gruba da aynı iki testler uygulanır. iki testten elde edilen puan kümesinin arasındaki korelasyon kat sayısı, genellikle Pearson çarpım moment formülü ile hesaplanarak eşdeğerlik katsayısı bulunur.

2. Test-tekrar test yöntemi: Testi alan bireyin farklı zamanlarda aynı teste nasıl bir tutarlılıkla yanıt vereceğini belirlemeye yarayan yöntemdir. Bireyin durumundaki geçici değişiklikler hataların ortaya çıkmasına sebep olmaktadır ve bu da bireyin gözlenen puanının gerçek puanı etrafında dalgalanmasına sebep olmaktadır. Aynı şekilde puanlama hataları,

yanlış işaretleme ve katılımcının davranışındaki geçici dalgalanmalar gözlemlenen puanlar üzerinde etkili olabilmektedir. Bu şekilde test uygulayıcısı bir testi uyguladıktan sonra katılımcının test maddelerini hatırlamayacakları kadar uzun ve yeni öğrenmeleri gerçekleştirmeyecekleri kadar kısa olmalıdır. Bu şekilde elde edilen iki puan kümesi arasında elde edilecek korelasyon ile kararlılık katsayısı elde edilebilir.

3. Karma yöntem: Test-tekrar test ve paralel formlar yönteminin birlikte kullanıldığı yöntemidir. İki eşdeğer formdan birinci ve ikinci formlar gruplara önce uygulanır, belirli bir süreden sonra formlar grubun yarısıyla değiştirilerek tekrar uygulanarak test puanları arasındaki güvenilirlik katsayısı olan eşdeğerlik ve kararlılık katsayısı elde edilir.

Paralel formlar, test-tekrar test ve karma yöntemler yukarıda ifade edildiği gibi birden fazla uygulama gerektiren yöntemlerdir. Ancak çoğu zaman bir testin birden fazla kez uygulanması; bireylerin veya testin ölçülen özellikleri, zaman ve ekonomik gibi nedenlerden dolayı gerçekleşmeyebilir. Böyle bir sınırlılığı ortadan kaldırmak için, paralel testler varsayımından yola çıkarak bir tek testin uygulamasından güvenilirlik kestirme yöntemleri ortaya konulmuştur. Burada tek bir test formunun maddeleri veya alt testlerinin bir gruba bir kez uygulanarak, testin kendisiyle olan tutarlılığına bakılır. Bu durumdaki güvenilirliği tahmin veya değerlendirmek için bulunan katsayı iç tutarlık katsayısı olarak ifade edilir. Uygulanan bir testin hatalı yönetilmesi ve puanlanması, tahmin etme ve test oturumundaki bireysel performanstaki geçici dalgalanmalardan dolayı ölçüm hataları tutarlılık katsayısını etkileyebilir. Ayrıca tek bir uygulamadan yola çıkarak iç tutarlık katsayısı değerlendirildiğinden dolayı hazırlanan testin aynı performansı ölçen maddelerden oluşması, maddelerin iyi yazılmış olması ve sınava girenlerin içerikle ilgisiz bir şekilde yanıt vermesine sebep olabilecek teknik unsurlardan arındırılmış olması gerekmektedir. Tek bir testin kullanılmasıyla güvenilirlik katsayısını tahmin etmek için kullanılan yöntemler; Yarıya Bölme yöntemi ve Madde kovaryanslarına dayanan yöntemler şeklinde gruplandırılmıştır (Crocker and Algina, 2006: 135-136; Erkuş vd., 2017: 25-26).

1. Yarıya bölme yöntemi: Bu yöntemle amaç bir kez uygulanan bir testin alt testlerinin paralel testler olduğu varsayımından yola çıkarak testi ikiye bölme ve bu işlem sırasında mümkün olduğunca paralel iki yarı test oluşturmaktır. Bu işlemde elde edilen iki yarı testin ortalama ve varyanslarının eşit olduğu kabul edilir. Testteki tek numaralı maddelerin 1. yarı teste, çift numaralı olan maddelerin 2. yarı teste atanması ile oluşan iki yarı testin uygun olacağı böylece testin başlangıç ve sonundaki yorgunluk, madde güçlükleri gibi hata kaynaklarının dengeleneceği kabul edilir. Yapılan işlemler sonucunda iki yarı test

arasındaki elde edilen korelasyon değeri arasında korelasyon katsayısı hesaplanır. İki yarı test, iki farklı testmiş gibi kabul edildiğinde Pearson Momentler Çarpımı Korelasyon Katsayısı ile korelasyon hesaplanır. Bu korelasyon katsayısı, testin iki yarısı için eşdeğerlik katsayısıdır. Bulunan değer, test yarıya bölüldüğünden bir alt testin güvenilirlik katsayısını verir. Bütün testin güvenilirlik katsayısı ise Spearman-Brown formülü ile bulunur. Spearman-Brown düzeltmesi formülü gerektirmeyen Rulon yöntemi, yarı testler arasındaki puan farklarının kullanımına dayanır. Böylece tüm testin güvenilirlik katsayısını verir. Guttman tekniğinde Rulon yönteminin eşitliğinden türetilerek bütün testin güvenilirliğini hesaplamaya çalışır(Erkuş vd., 2017: 26-28; Crocker and Algina, 2006: 136-138).

2. Madde kovaryansına dayanan yöntemler: Madde kovaryansına dayanan güvenilirlik hesaplama yöntemleri de paralel testler varsayımı üzerine temellenmektedir. Bu yöntemlerde test maddelerinin her birinin kendi arasında paralel testler oluşturduğu ve test maddelerinin paralel testin bir bileşkesi olduğu kabul edilir. Bu varsayımla birlikte aynı ortalama ve varyansa sahip olma varsayımından yola çıkarak ikili puanlanan maddeler için KR-20 ve bununla birlikte bütün maddelerin güçlüğünün eşit olduğu durumlarda ise KR-21 yöntemleri kullanılmaktadır. Madde güçlüğünün değişmesi durumunda KR-21'den elde edilen güvenilirlik katsayısı KR-20'den daha düşük çıkacaktır. Bir iç tutarlık katsayısı olan Cronbach Alfa (α) yöntemi ise çoklu puanlanan maddelere uygulanan bir yöntemdir. Bilgisayarla güvenilirlik hesaplanmasına imkan sağlayan ve kullanışlı paket programı olan Hoyt'un (1941) varyans analizi yöntemi, bireyler ile maddeler arasındaki değişimi gözleme imkanı sağlamaktadır(Crocker and Algina, 2006: 138-141; Erkuş vd., 2017: 28-34).

Hambleton and Swaminathan, (1985: 2-3) göre KTK'nın varsayımları incelendiğinde bazı eksikliklere rastlandığını ve KTK'nın zayıf bir model olduğu ifade edilmektedir. Sınırlılıklar şu şekilde ifade edilmiştir:

1. Test maddelerinden elde edilen madde güçlüğü ve madde ayırt edicilik değerleri, testi cevaplayan grubun yeteneğine bağlı olmalıdır. Testi cevaplayanların ortalama yetenek seviyesi ve yetenek puanlarının aralığı, genellikle önemli ölçüde, madde istatistiklerinin değerlerini etkiler.

2. Güvenilirlik kavramının paralel testler varsayımı açısından tanımlanmasından dolayı pratikte paralel testler elde etmek zordur. Bu nedenle paralel test olarak kabul edilen iki test karşılaştırıldığında, zorluk düzeyinde önemli ölçüde farklılık gösteren iki testte yüzde 50 oranında performans gösteren iki sınav, yetenek açısından eşdeğer kabul edilemez.

3. Kaynağının belli olmadığı tesadüfi hatalardan dolayı bireylerin iki testteki durumu aynı değildir.

4. KTK'nın, bir bireyin bir test maddesi ile karşılaştığında bu maddeye karşı nasıl bir performans sergileyeceğini belirlemek için herhangi bir temeli bulunmamaktadır. Sınava giren kişinin herhangi bir soruyu doğru cevaplayabilme olasılığını bilmek, bir testi hazırlarken bireyin yetenek düzeyi ile test maddelerinin eşleşmesi bakımından önemlidir.

5. Ölçme hata varyanslarının sınava giren tüm bireyler için aynı kabul edilmesi KTK'nın başka bir sınırlılığıdır.

Bunların dışında KTK; test tasarımı, yanlış maddelerin belirlenmesi, test puanlarının hesaplanması, madde ayırt edicilik gücü gibi bilgileri vermekte ve bireylerin gerçek yeteneklerinin ele alınması açısından yetersiz kalmaktadır.

2.2.2. Madde Tepki Kuramı (MTK)

Testler genel olarak maddelerden oluşur ve test puanı da madde puanlarının toplamından oluşmaktadır. KTK, elde edilen bu test puanlarından yola çıkarak istatistik çıkarımlar yapmaktadır. Ancak bu teori test uygulayıcısının kontrolü dışındaki konular hakkında bir bilgi sunmamaktadır. Test maddeleri daha önce benzer kişilere uygulanmamışsa, bireylerin maddelere nasıl tepki vereceğini tahmin etmemiz zordur. Pratik bir test geliştirme çalışmasında, sınava giren herhangi bir gruba test uygulandığında hazırlanan herhangi bir testin istatistiksel ve psikometrik özelliklerini tahmin edebilmemiz gerekmektedir. Test maddeleri daha önce katılımcılara uygulanmamışsa katılımcıların bu maddelere nasıl tepkiler verebilecekleri tahmin edilememektedir. Madde parametrelerini inceleyerek bir bireyin daha önce o madde ile karşılaşmamış olsa bile maddeye verebileceği tepkiyi olasılıksal olarak bilmemiz gerekmektedir. Bu sayede bireyin maddeye verebileceği tepki ile maddelerin hangi yetenek düzeyinde ölçümler için kullanılabileceğini tahmin edebiliriz. Ancak KTK bunların hiçbirini sunmamaktadır. Madde Tepki Kuramı (MTK) ise testi alan bireyin maddeye yanıtının yetenek veya beceri düzeyi ile nasıl ilişkili olduğunu matematiksel olarak ortaya koymaktadır. Bu ilişki Madde Karakteristik Eğrisi ile ortaya konulmaktadır (Baker, 2016: 2-3; Lord, 1980: 11-12).

Bir test maddesini cevaplayan katılımcının maddenin altında yatan özelliğe bir miktar sahip olduğu varsayılır. Buna bakılarak her katılımcının yetenek ölçüğü üzerinde yer alan bir

puana sahip olduğu söylenebilir. Yetenek ölçeği, “theta- Θ ” ile gösterilmektedir. Yetenek düzeyleri incelendiğinde bu yetenek düzeyindeki katılımcıların maddeleri cevaplamalarına ait belirli bir olasılık bulunmaktadır. Bu olasılık $P(\Theta)$ şeklinde gösterilir. Bir madde üzerinden düşünüldüğünde düşük yetenek düzeyine sahip bireylerin maddeyi doğru cevaplama olasılık değeri küçük iken yüksek yetenek düzeyine sahip bireylerin maddeyi doğru cevaplama olasılık değeri de büyüktür(Baker, 2016: 3).

MTK’da Madde Karakteristik Eğrisi’nin (MKE) önemli bir yeri vardır. Çünkü MTK’daki bütün yapılar bu eğri üzerine kurulmuştur. MKE üzerinde dikey eksen bir maddenin doğru cevaplanma olasılığını ifade ederken yatay eksen bireyin yetenek düzeyini ifade etmektedir. Bireyin yetenek düzeyi arttıkça bir maddenin doğru cevaplanabilme olasılığı da artmaktadır(Erkuş vd. 2017: 124).

MKE’i tanımlamak için kullanılan iki teknik vardır: Madde güçlüğü (b) ve madde ayırt ediciliği (a) dir. Madde güçlüğü, maddenin yetenek ölçeğindeki hangi noktada işlevsel olduğunu ifade eder. Bir maddenin kolay olması düşük bir yetenek düzeyindeki bireyler için daha kullanışlı iken zor bir madde ise yüksek yetenekli bireyler için daha kullanışlıdır. B parametresi maddenin hangi yetenek seviyesindeki bireylerin %50 olasılıkla doğru cevaplayabileceğini ifade eder. Madde ayırt ediciliği ise maddenin bulunduğu noktanın üstündeki ve altındaki yeteneğe sahip bireylerin birbirinden ne kadar ayırt ettiğini gösterir. Diğer bir ifadeyle MKE üzerinde orta noktadaki dikliği ifade eder. MKE üzerinde daha dik eğriler, maddenin ayırt ediciliğinin yüksek olduğunu ifade ederken daha düz eğriler madde ayırt ediciliğinin düşük olduğunu gösterir(Baker, 2016: 4).

Madde güçlüğü ve madde ayırt ediciliği parametreleri dışında maddeyi doğru cevaplayabilme yeteneği olmadığı halde bireyin şans faktörüne bağlı olarak maddeyi doğru cevaplayabilme olasılığının sürece dâhil edildiği c parametresi de bulunmaktadır(Erkuş vd. 2017: 133).

MTK’da doğru kestirimlerde bulunabilmek için temel anlamda bazı varsayımların sağlanması gerekmektedir. MTK varsayımları Demars (2016: 37-51) tarafından şu şekilde ifade edilmektedir:

- Tek boyutluluk: Bir testin tek bir yapıyı ölçen maddelerden oluştuğunu ifade eder. Modelde her bir bireyin tek bir theta değerinin olduğu kabul edilir. Maddeye verilen yanıt etkileyen faktörlerin sadece o maddeye özgü random hatalardan olduğu kabul edilir. Geriye kalan maddeler bu durumdan etkilenmez. Bu varsayımın ihmal edilmesi standart hata veya parametrelerin yanlış kestirmesine sebep olabilir. Çok boyutlu MTK modelleri bunun dışında tutulmaktadır.

- Yerel bağımsızlık: Bu varsayım maddelerin belirli bir yetenek düzeyine göre bağımsız olmasını ifade eder.

- Uyum (uygun veri modeli): Model ile verinin uyuşmamasından dolayı yanlış hesaplamalar yapılabilir. Bu nedenle belirlenen model doğru bilgiler sunmayabilir.

Yukarıda ifade edildiği gibi bir veriden doğru bilgiler elde edebilmek için model-veri uyumunun önemli bir yeri vardır. Bu nedenle bazı zamanla MTK bazı modeller ortaya konulmuştur. Bu modeller 1, 2 ve 3 parametrelili lojistik modeldir.

1 parametrelili lojistik model (1PL) ve rasch modeli tek bir parametre kullanılarak maddeler üzerinde bireyin başarısını olma olasılığını kestirmek için kullanılan modeldir. 1PL modelde tüm maddelerde a parametresi aynı değeri alırken rasch modelinde 1 değerini alır ve c değeri de sifira eşitlenecek şekilde sabitlenir.

2 parametrelili lojistik model (2PL), Lord (1952) tarafından normal dağılım eğrisine göre ilk olarak geliştirilen bu model daha sonra madde ve yetenek parametrelerine bağlı integralleme işleminin yarattığı sorunlardan dolayı, matematiksel ve istatistiksel açıdan avantaj sağladığı için lojistik fonksiyonla kullanılmıştır. 2PL'de a ve b parametreleri kullanılarak madde karakteristik eğrisi yorumlanır.

3 parametrelili lojistik model (3PL) ise a ve b parametrelerine ek olarak c parametresinin dâhil edildiği modeldir. Modelde maddeyi şansla doğru cevaplayabilme olasılığını "c" ile karakterize ederek iki parametrelili modele üçüncü bir parametre eklenmiştir. c parametresinin sıfır (0) olması durumunda 3PL model 2PL modele dönüşür. c parametresinde dolayı madde karakteristik eğrisinde düşük asimptot, yukarı doğru çıkar. Bu sebeple c parametresi madde karakteristik eğrisini kısmen de olsa dikey eksende ötelediği ifade edilebilir. böyle bir durumda b parametresi ile 0 aynı yetenek düzeyindeki bireylerin maddeyi doğru yanıtlama olasılıkları .50 den daha büyük olacağı söylenebilir(Baker, 2016: 18-26; Demars, 2016: 11-20; Erkuş vd. 2017: 131-135).

2.2.3. Genellenebilirlik Kuramı

Bilimsel araştırmaların amacı nesnelere özellikleri veya nitelikleri hakkında bilgi sahibi olmaktır. Bu nedenle özellikler veya nitelikler hakkında bilgi toplarken ölçme kurallarına uymak önem taşımaktadır. Herhangi bir ölçme işlemi esnasında ölçme sonuçlarına hatalar karışabilmektedir. Ölçme sonuçlarına hata karışması bütün bilimsel araştırmaların temel sorunu olarak karşımıza çıkmaktadır. Eğitim ve psikoloji alanında bu tür sorunlar güvenilirlik adı altında karşımıza çıkmaktadır. Genel olarak incelendiğinde güvenilirlik

gözlemlenen puanlardaki tutarlılık veya tutarsızlıkların ölçülmesi olarak ifade edilmektedir. Genellenebilirlik kuramı (G Kuramı), bir ölçme işleminde elde edilen gözlenen puanlarda ortaya çıkan bu tutarsızlıkların kaynaklarının ölçülerek belirlenmesini ve ayırt edilmesini sağlayan bir kuram olarak karşımıza çıkmaktadır. Bu yönüyle KTK'nın bir uzantısı olarak görülmektedir(Brennan, 2011: 1-2).

G kuramının KTK'nın bir uzantısı olarak görülmesinin nedeni, sayısız ölçme için yeterince güçlü bir istatistiksel yoldan biri olan varyans analizini (ANOVA) kullanarak kapsamlı bir kavramsal temel oluşturmasıdır. KTK'nın temel denklemi $X=T+E$ olarak ifade edilmektedir. KTK da güvenilirlik, bir ölçme işleminde tek bir gerçek puana ait her bir gözlem veya test puanının paralel gözlemlerin bir grubuna ait tek bir güvenilirlik katsayısını ortaya koyması fikrine dayanır(Lord and Novick, 1968: 28-29; Gulliksen, 1950: 4-5). Bu denklem paralel formlar dikkatli incelenip eşitlendiğinde uygunken, puanlayıcıların varyansları ve merkezi eğilimleri farklı olduğunda, gözlemlerden meydana gelen kapsama bağlı olarak ve yapılar heterojen olduğunda gerçekçi olmaz ve kısıtlayıcıdır (Atılğan, 2019: 2).

G kuramının KTK'dan üstünlüğü, tek bir analiz yardımıyla farklı hata kaynaklarının aynı anda kestirilebilmesine olanak sağlamasıdır. Bir araştırmacının yapmış olduğu ölçme işleminde birden çok hata kaynağı bulunuyorsa (çalışmada yer alan hata kaynaklarının sayısı K olmak üzere) G kuramının dayandığı model;

$$X= \mu_p +E_1 + E_2 + \dots +E_K$$

şeklinde gösterilmektedir. Eşitlikte yer alan “ μ_p ” KTK'daki gerçek puan olarak yorumlanan evren puanı (T)'dir (Brennan, 2011: 7). G kuramı; KTK'ndaki Spearman Brown formülünü kullanarak madde sayısının güvenilirliğe etkisini kestirmenin daha ötesinde, güvenilir sonuçlar elde etmek amacıyla gereksinim duyulan durum, test formu ve puanlayıcı gibi olası bütün hata kaynakları sayısının ne olması gerektiğini belirlemede yardımcı olur. G kuramı, KTK'nında elde edilen güvenilirlik katsayısına benzer olan ve güvenilirlik derecesini belirten genellenebilirlik katsayısını verir(Güler, Kaya Uyanık ve Taşdelen Teker, 2012: 5).

KTK, ölçme hatalarını tesadüfi hata olarak ele almaktadır ve farklı değişkenlik kaynaklarından gelen hataları potansiyel hata kaynaklarından ayırmamaktadır. Ancak G kuramı sistematik veya sistematik olmayan hata kaynaklarını ele alarak aynı anda hesaplamayı sağlar. G kuramının bu yönüyle güvenilirlik kestiriminin avantajı, tüm değişkenlik

kaynaklarını ele alarak, bir tek hesaplama ile kapsamlı bir güvenilirlik elde edilmesini sağlamasıdır. G kuramının yardımıyla bir tek çalışmada bütün potansiyel hata kaynaklarının tek tek veya birbirleriyle etkileşiminden ortaya çıkan hataları değerlendirerek bir kat sayının elde edilmesini sağlar (Atılğan, 2004: 15).

G kuramında potansiyel ölçme hataların kaynakları (madde, puanlayıcı, form, zaman vb.) “yüzey” (facet) olarak isimlendirilir. İfade edilen bu hata kaynakları ölçmedeki olası hata kaynakları olarak ele alınmaktadır. Puanlayıcının değişkenlik kaynağı olarak ele alındığı bir çalışmada her bir puanlayıcı koşul olarak ele alınmaktadır. Bu tanım her bir madde, tekrarlı ölçüm vb. içinde olabilmektedir. G Kuramında kullanılan yüzey ve koşul ifadeleri geleneksel varyans analizindeki faktör (factor) ve düzey (level) kavramlarına denk gelmektedir (Atılğan, 2019: 5; Crocker and Algina, 2006: 159; Güler vd., 2012: 6-7).

Yapılan bir G çalışmasında araştırmacı, üzerinde ölçme yaptığı örnekleminin ölçme evreninin genellemesiyle ilgilenmektedir. *Evren*; çoğunlukla ölçümlerin elde edildiği ve örneklemden daha kapsamlı olan ölçme durumları şeklinde tanımlanmaktadır. Bir G çalışması belirli koşullar altında elde edilen ölçümlere dayanmaktadır. Belirlenen bu koşullar çoğunlukla daha geniş setteki koşulların birer temsilcisi konumundadır. G çalışmalarında bulunan yüzeylere bağlı koşullardan elde edilen ölçümlere “*genelleme evreni*” denir. Diğer bir ifadeyle genelleme evreni, araştırmayı yapan kişinin genellemek istediği koşullar seti olarak ele alınabilir. Araştırma sonucunda ele alınabilecek bütün koşullardan elde edilen ölçme sonuçları evrenine de “*kabul edilebilir gözlemler evreni*” denir. G çalışması sonucunda elde edilen ölçmeler sonucunda kabul edilebilir gözlemler evreni genelleme evrenini kapsamı şartıyla, yapılacak bir K çalışmasının planlanmasına zemin hazırlanacaktır. Bir çalışmadaki kabul edilebilir gözlemler evreni o çalışmada yer almayan farklı yüzeyleri barındırmadığından, yapılması istenen K çalışması da sadece çalışmada yer alan yüzeylere bağlı olarak kurulabilmektedir. Bir G çalışmasında bulunmayan farklı yüzeylerden K çalışması kurmak mümkün görünmemektedir (Brennan, 2001: 9; Güler vd., 2012: 5-7).

Evren puanı, kabul edilebilir gözlemler evreninden elde edilen puanların ortalaması olarak kabul edilir. Amaçlanan değişkenlik kaynakları üzerinden elde edilen bu puan ölçme puanı olarak kabul edilmektedir. Elde edilen evren puanı varyansı da KTK”da elde edilen gerçek puan varyansına benzemektedir (Erkuş vd. 2017: 217).

G kuramı ile yapılan çalışmalarda amacına göre genellenebilirlik (G) çalışması ve karar (K veya D) çalışması şeklinde iki farklı çalışma yapılabilmektedir. G çalışmasının amacı üzerinde ölçme çalışması yapılan örneklemin ölçme evrenine genellenmesidir. Bir testten alınan yanıtların farklı zamanlarda değişmemesi, diğer bir ifadeyle kararlılığı, birden fazla form (test) arasındaki puanların tutarlılığı, alt testler veya maddeler arası iç tutarlılığın araştırılması G çalışması kapsamında incelenmektedir. G çalışması ölçme çalışmalarında ortaya varyans kaynakları hakkında bilgi sağlayarak ölçme hakkında bilgi sunar ve bu şekilde ölçme araçlarının geliştirilmesi ya da daha sonraki kullanımlarda değişkenlik kaynaklarından ortaya çıkan hata kaynaklarının belirlenerek azaltılmasına yardımcı olmaktadır. Öte yandan K çalışması ise G çalışmasından elde edilen bilgilerden yararlanarak ölçme işlemlerindeki hataların en aza indirilmesinin yollarını araştırarak belirli bir amaç için yapılan ölçmenin en uygun deseninin belirlenmesini sağlamaktır (Güler vd., 2012: 5-6). K çalışmaları ölçme nesnesi veya kişisi hakkında önemli kararlar verilmesi açısından önemlidir. Diğer bir ifadeyle kişi üzerinde tahminde bulunma veya yorumda bulunmayı sağlamaktadır(Brennan, 2001: 8).

Bir araştırmada, belirlenen yüzeyler araştırmacının kararına göre rastgele veya sabit yüzey şeklinde seçilebilir. Rastgele (random) yüzeyler, örneklemin dışındaki evrene genelleme yapmanın önünü açar. Bu sebeple yüzeylerin koşuluna bağlı olarak seçilen örnekleme ilgilidir. Seçilen bir örneklemin tesadüfi örneklem olma şartı, örneklem büyüklüğünün evren büyüklüğünden oldukça küçük olması, seçilen örneklemin belirlenen evrenden rastgele çekilmesi veya aynı evren üzerinden seçilen benzer büyüklükteki örneklemlerle değiştirilebilmesi gerekmektedir. Sabit (fixed) yüzeyde ise evrenin küçük olması nedeniyle evren üzerinde çalışılması veya genellemelerin uygun olmaması sebebiyle kasıtlı olarak seçilen koşullar dışında seçilen ve bunun dışında genellemeyle ilgilenilmeyen durumlarda kullanılır. Diğer bir ifadeyle rastgele yüzeylerde evrene genelleme yapılırken, sabit yüzeylerde örneklem üzerinde genelleme yapılmaktadır(Atılğan, 2019: 9; Erkuş vd. 2017: 217; Shavelson and Webb, 1991: 12).

G kuramında veri yapısına göre çaprazlanmış (crossed) veya yuvalanmış (nested) şeklinde iki tür desen kullanılmaktadır. Bir ölçme işleminde bir değişkenlik kaynağında bulunan tüm koşulların öteki değişkenlik kaynağındaki bütün koşullarını gözlemlemesi sonucu çaprazlanmış desen ortaya çıkar. Böyle bir durumda değişkenlik kaynaklarının arasına "x" işareti konulmaktadır. Eğer bir değişkenlik kaynağında bulunan bütün koşullar diğer değişkenlik kaynağında bulunan koşullar tarafından gözlemlenmiyorsa diğer bir deyişle belirli değişkenlik kaynağının koşulları tarafından gözlemleniyorsa yuvalanmış desen kullanılmış

olur. Yuvalanmış desenlerde değişkenlik kaynakları arasına ise “:” işareti konur(Erkuş vd. 2017: 218; Shavelson and Webb, 1991: 22).

10 adet benzer ürünün 2 puanlayıcı tarafından puanlandığı düşünüldüğünde, her bir puanlayıcının bütün ürünleri puanlaması durumunda çaprazlanmış desen, her bit puanlayıcının belirli ürünü puanlaması durumunda ise yuvalanmış desen kullanılmış olmaktadır.

G kuramına göre ölçme, kabul edilebilir gözlemler evreninin bir örnekleme şeklindedir. Yapılan gözlemlerde verilecek bir kararın amacına göre değişkenlik gösterebilmektedir. Bu nedenle yapılan çalışmalar Tek yüzeyle evrenler, yüzeyle evrenler veya üç veya daha çok yüzeyle evrenlere göre yapılmaktadır. Tek yüzeyle evrende sadece bir hata kaynağı ortaya çıkar. Yapılan bir araştırmada araştırmacı ölçme aracında kullandığı test maddelerini daha geniş bir madde grubuna genellemeyi istiyorsa, testteki maddeler ölçmenin değişkenlik kaynağı olarak ele alınır. Kabul edilebilir olan bütün maddeler ise madde evreni olarak kabul edilir. Bir çalışmada hangi değişken genellenmek istenirse istensin, herhangi bir ölçme örnekleme istenilen evrene genellenmek istendiğinde hatanın ortaya çıkması kaçınılmazdır(Güler vd., 2012: 11).

Bir evrende bulunan bütün maddelerin güçlükleri aynı ve bireyin aldığı puan maddeden maddeye yaklaşık aynı değeri aynıysa, böyle bir durumda bireyin bir madde üzerindeki performansı diğer tüm maddelere genellenebilme olanağı sunar. Fakat testteki madde güçlüklerinin farklı olması durumunda, birey aldığı puan teste bulunan madde örneklemine göre olacaktır. Bu durumda ise örneklemden evrene genelleme yapma riskli olur ve maddelerdeki bu değişkenlik, genellemenin potansiyel hata kaynağı olur.

Bir çalışmada bireyin evren puanına ulaşılması istenir fakat bütün maddelerin bireye uygulanması sonucunda böyle bir durum mümkün olabilmektedir. İstenilen bu durumda pratikte mümkün olmadığından, istenilen duruma ait verilere ulaşması da mümkün değildir. Fakat araştırmayı yapan kişinin amacı öğrencinin başarılarının göstergesi için kullanılan madde örnekleminin, kabul edilebilir bütün maddelerin oluşturmuş olduğu evrene ne derecede doğru genellenebildiğini kestirmeye çalışmaktır.

Tek yüzeyle desende dört farklı değişkenlik kaynağı ortaya çıkar. Birinci değişkenlik kaynağı bireylerin başarılarındaki farklılıktan doğar. Bu değişkenlik kaynağı “ölçme objesi” olarak da adlandırılır. Ölçme objesindeki değişkenlik bireyin yetenek, bilgi, tutum gibi değişkenliklerini ifade eder.

İkinci deęişkenlik kaynaęı, maddelerin güçlüklerinin farklılıęından ortaya çıkar. Maddelerin kolay, zor veya orta güçlüktedir. Madde güçlüklerindeki bu farklılıklar madde örnekleminin madde evrenine genellenmesini güçleştirmektedir.

Üçüncü deęişkenlik kaynaęı, testi yanıtlayan bireyin geçmiş yaşantıları; bilgi ve deneyimleridir. Bireylerin daha önce karşılaştıkları konular hakkındaki maddeleri doğru yanıtlayma olasılıkları daha yüksektir. Bireyin yaşantıları sebebiyle bir maddeden aldığı başarı puanı eşleşmesi deęişkenlięi artırır. Bu durumda genellenmenin güçleşmesine sebep olur.

Dördüncü deęişkenlik kaynaęı ise kaynaęı bilinmeyen veya kaynaęı tanımlanamayan deęişkenlerdir.

Bir başarı testindeki puanların varyanslar kaynakları özetle;

- 1- Ölçme objelerinden doğan farklılıklar
- 2- Madde güçlükleri arasındaki farklılıklar
- 3- Birey ile madde arasındaki etkileşim
- 4- Kaynaęı bilinmeyen veya tanımlanamayan durumlar

Madde ile birey arasındaki etkileşim ile kaynaęı bilinmeyen deęişkenlik kaynakları ayrıştırılmadığında dolayı bu iki deęişkenlik kaynaęı birlikte *artık (residual)* olarak kabul edilir.

Tablo 1. Tek yüzeyli ölçmeler için varyans kaynakları

<i>Değişkenlik kaynağı</i>	<i>Değişkenlik türü</i>	<i>Varyans sembolü</i>
Birey (b)	Evren puanı	σ_b^2
Madde (m)	Koşullar	σ_m^2
bxm etkileşimi ve tanımlanamayan (e)	Artık	$\sigma_{b,m,e}^2$

G kuramı, **tablo 1**'de bulunan bu üç varyansın büyüklüğünün kestirilmesini sağlar ve her bir değişkenlik kaynağının büyüklüğünü varyans bileşeni şeklinde ifade eder. G kuramında amaç, önemli değişkenlik kaynaklarını tespit etmek ve bireye ait değişkenlik dışındaki değişkenliklerin daha sonra yapılacak başka bir araştırma için en aza indirilmesini sağlamaktır. Bu uygulama sayesinde yapılan ölçmenin yeniden tasarlanmasına olanak sağlanmış olur. **Tablo 1** incelendiğinde bu ölçme işleminde bir tek yüzey bulunmaktadır. Bu yüzey madde yüzeyidir.

G çalışmaları için tek yüzeyli evrende yapılabilecek iki farklı desen söz konusudur. "b" birey ve "m" madde olmak üzere bunlar; bxm (çaprazlanmış desen) ve b:m (yuvalanmış desen) desenleridir. Aynı şekilde K çalışmaları içinde bxM (çaprazlanmış desen) ve b:M (yuvalanmış desen) desenleri kullanılır.

Ölçme çalışmalarda bazı durumlar iki yüzey bulunabilir. Diğer bir deyişle yapılan ölçmelerde iki farklı değişkenlik kaynağından hatanın karışması söz konusu olabilmektedir. Bu durumda yapılan çalışmada iki yüzeyli evrende ölçme objesinin dışında altı farklı değişkenlik kaynağı ortaya çıkar.

Tablo 2. İki yüzeyli ölçmeler için varyans kaynakları

<i>Değişkenlik kaynağı</i>	<i>Değişkenlik türü</i>	<i>Varyans sembolü</i>
Birey (b)	Evren puanı	σ_b^2
Madde (m)	Maddeden maddeye değişip bireyin davranışlarından kaynaklanan tüm bireyler üzerindeki sabit etki	σ_m^2
Puanlayıcı (p)	Puanlayıcılardan kaynaklanan bütün bireyler üzerindeki sabit etki	σ_p^2
bxm	Maddeden maddeye bireylerin yanıtlarındaki farklılıklar	σ_{bm}^2
Bxp	Puanlama esnasında puanlayıcıları arası tutarsızlık	σ_{bp}^2
Mxp	Maddeden maddeye puanlayıcı sıklığı arasındaki farkın sebep olduğu sabit etki	σ_{mp}^2
bxm _p , e	Artık/ kalan varyans	$\sigma_{bmp,e}^2$

Tablo 2 'de görüldüğü gibi ölçme objesi dışında kalan altı değişkenlik kaynağı da örneklemin evrene genellenmesi durumunda hatalara sebep olacaktır. Madde güçlüklerindeki farklılıklar, puanlayıcılar arası tutarsızlık, birey ile madde etkileşiminde eski yaşantıların etkisi, puanlayıcıların bireylere karşı katı veya cömert davranması, puanlayıcının maddeye karşı tutumu ve kaynağı belli olmayan değişkenlik kaynaklarının tümü iki yüzeyli desenlerde karşımıza çıkabilmektedir.

Sosyal bilimlerdeki ölçme işlemlerinin karmaşıklığı nedeniyle ikiden fazla yüzey kullanılabilir. Fakat kabul edilebilir gözlemler evreni genişliği artıka, örneklemden evrene yapılan genellemelerde hata yapılma olasılığı artacaktır(Güler vd., 2012: 18).

G kuramının avantajı elde edilen puanların güvenilirliğinin farklı yorumlamalar için kestirilebilmesini sağlamasıdır. KTK da elde edilen sonuçlara göre yapılabilecek tek yorum, bireylerin birbirleri arasında durumlarını karşılaştırmaya olanak sağlayarak bağıl değerlendirme yapmasıdır. Ancak G kuramı araştırmacılar için hem bağıl değerlendirme için hem de mutlak değerlendirmeye olanak sağlar. G kuramında bağıl me mutlak değerlendirme genellenebilirlik (generalizability) katsayısı (G veya Ep^2) ve güvenilirlik katsayısı (Φ veya “phi”) şeklinde iki türlü hesaplanmaktadır(Güler vd., 2012: 20; Shavelson and Webb, 1991: 92-93).

Bağıl değerlendirmelerde kullanılan G katsayısı sosyal bilimlerde yapılan ölçmelerde örneklemden elde edilen gözlenen puanların evren puanına hangi doğrulukta genellenebileceğini görmemizi sağlar. Genellenebilirlik katsayısı, birey varyansı σ_b^2 ve bağıl hata varyansı σ_δ^2 olmak üzere;

$$Ep^2 = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_\delta^2}$$

formülüyle bulunur. G katsayısının değeri 0-1 aralığındadır. 0,80 üzeri değer G katsayısı için kabul edilebilir bir değerdir.

Mutlak değerlendirme için kullanılan güvenilirlik katsayısı olan Φ “phi” katsayısı bir yüzeyli rastgele desenlerde KTK’daki Cronbach Alpha değerine benzer sonuçlar vermektedir. 0-1 aralığında değer alan Φ “phi” katsayısı, birey varyansı σ_b^2 ve mutlak hata varyansı σ_Δ^2 olmak üzere;

$$\Phi = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_\Delta^2}$$

formülüyle gösterilir (Cardinet, Johnson and Pini, 2010: 28-29).

2.3. Performans ve Ürünlerin Ölçülmesi

Performans ölçümü, bir öğrencinin ürün oluşturmasını veya davranışı ortaya koymasını zorunlu kılarak açıkça belirlenmiş ölçütlere bağlı kalarak öğrencinin çalışmalarının niteliğini değerlendirmek amacıyla kullanılır. Bir performans ölçümünde bireylerin kendi bilgilerini kullanarak bir şeyler ortaya koymasını gerektirir. Bunlar bir ürün olabileceği gibi bir davranış veya bir proje çalışması da olabilmektedir(Nikto and Brookhart, 2016: 246).

Bir performans ölçümü hem performans görevinin kendisi hem de puanlama için kullanılacak puanlama ölçeği olmak üzere iki bölümden oluşmaktadır. Puanlama ölçeği, daha önce belirlenen ölçütlere uygun olarak hazırlanmalıdır. Puanlama ölçeklerinin kullanılmadığı sınıflarda öğretim etkinlikleri performans ölçümü olarak değerlendirilmemektedir. Uygulamalar esnasında bu durumlar bulanıklaşır. İyi öğretmenler sınıflarında sürekli olarak öğrencilerin çalışmalarını gözlemlemektedir. Öğretmenin bu gözlemlerde aklında bulunan bazı ölçütler biçimlendirici amaçlar için kullanabilmektedir. Öğretim sırasında Öğrencilerin ilgileriyle deneyimler elde etmeleri ile beraber öğrenme amaçlarına bağlı olarak birçok uygulama yapmaları gerekmektedir. Bu sebeple görev ve ölçütler birlikte bütünleyici ve notlandırılmış performans ölçümlerini zorunlu kılmaktadır(Nikto and Brookhart, 2016: 246).

Öğrencinin ortaya koyduğu performans, herhangi bir ürünü ortaya koymakla sonuçlandığında genellikle süreç veya yöntem yerine ürünün yargılanıp derecelendirmesi istenir. Öğretmenler ilk aşamada yöntemi değerlendirebileceği gibi temel becerilerin oluşmasından sonra ikinci aşamada ürünü değerlendirebilir. Bir ürünün değerlendirilmesi birçok alanda ölçme bilgisi gerektirmektedir. Bu açıdan dereceleme ölçekleri ürün ölçmelerde bir dereceye kadar aynı amaca hizmet etmektedir. Bu sebeple aynı özellikler anlamında bütün öğrencilerin ürünlerini değerlendirmemize yardımcı olmakta ve ürünün kalitesine önem verilmektedir(Tekindal, 2017: 258).

Yeni öğretim programları incelendiğinde bazı ölçme ve değerlendirme araçları önerilmiştir. Önerilen bu ölçme ve değerlendirme araçlarının etkinlikler üzerinde işe yarayabilmesi için bazı koşulların karşılanması gerekmektedir. Bunlar; gözlenen veya incelenen süreç veya ürünün bireyin kendi performansı veya ürününün olduğunun belirlenmesidir. Gözlemler ve incelemelerin uygun koşullarda yapılması ve doğru bir şekilde sayısallaştırılması da bir diğer önemli konudur. Bu şekilde düşünüldüğünde yapılan çalışmaların öğretmenin denetiminde yapılması ortaya çıkan ürünün veya performansın öğrenciye ait olduğu sağlanabilir. Bu sayede yapılan gözlemler ve değerlendirmeler ile birlikte öğrencinin öğrenme eksiklikleri veya öğrenme düzeyi belirlenebilir(Özçelik, 2013: 77-78).

Okulların en önemli görevi öğrencilerin hangi konularda eksikliklerinin olduğunu hangi konularda başarılı olduğunu gösterebilmesidir. Öğrencilerin eksikliklerini bilmesi veya başarılı yönlerini bilmesi böyle bir durumda önem arz etmektedir. Çünkü birçok öğrenci öğretmenlerinin kendilerini sevmediği için düşük not aldığını veya belirli bir alanda çok iyi olduğu için yüksek not aldığını düşünmektedir(Kutlu vd., 2017: 51).

Bu bilgilerden yola çıkarak bir öğrencinin ortaya koymuş olduğu performans veya ürünün puanlanması esnasında nesnel puanlamalar yapabilmek için dereceli puanlama puanlarının önemli bir yeri bulunmaktadır.

2.3.1. Dereceli Puanlama Anahtarı (Rubrik)

Tamamlayıcı değerlendirme tekniklerinde ya hep ya hiç şeklindeki kurallar ölçüt alınmaz. Diğer bir ifadeyle başarı ve başarısızlık arasında bireyin çalışmalarında kısmi olarak performansını ve bilgisini de değerlendirmeye çalışır. Bu amaçla puanlama anahtarları (rubrikler) kullanılır. Puanlama anahtarı, bir bireyin bir kavram, durum veya olayla ilgili bilgisini belirlemek veya bir çalışmaya ilişkin yeterliliğini belirlemek için kullanılan bir puanlama sistemidir. Buna ek olarak puanlama anahtarları, öğrencilerin zayıf ve güçlü yönleri ile ilgili geri bildirim verme, iyileştirme ve düzenleme amaçlı öğretimi planlama gibi nedenlerle de kullanılabilir(Bahar vd., 2015: 50).

Dereceli puanlama anahtarının yapısı diğer klasik puanlama anahtarlarının yapısından farklıdır. Puanlama anahtarlarının oluştururken şu temel elemanlar göz önüne alınabilir;

1. Bireyin belirli bir görev veya çalışmada göstermiş olduğu başarı hakkındaki durumunu belirlemede kullanılacak bir veya birden fazla ölçüt.

2. Belirlenen ölçütlerin ne anlama geldiğini ifade eden betimsel anlatımlar.

3. Her bir ölçüt düzeyinde bireyin başarısını gösterecek düzey veya dereceler.

4. Uygunsa her bir düzeyin hangi anlamda geldiğini ifade eden betimsel açıklamalar.

Bu ifadelerden yola çıkarak dereceli puanlama anahtarındaki ölçütler sayesinde puanlayıcıdan puanlayıcıya değişmeyen standart ve nesnel bir belirleme yapma, öğrencinin durumu hakkında karar vererek açık ve anlaşılır bir not verme işlemi yapılmasına katkı sağlanabilir (Kutlu vd., 2017: 53).

Bir puanlama ölçeği öğrencinin performansının niteliğini değerlendirmek için kullanıldığından dolayı bir dizi uyumlu kurallar barındırmaktadır. Bu kurallar karar verme sürecinde puanlayıcıları yönlendirir ve karar verme sürecinde kararlılıklarını sağlar. Bir derecelendirme ölçeğinde bireylerin performanslarının niteliğinin seviyesini yansıtmak için 0-

3, 1-4 aralığındaki sayıları içermektedir. Burada ifade edilen her sayı ortaya konulan bir performansın niteliğinin sözel olarak karşılığını temsil etmektedir(Nikto and Brookhart, 2016: 248).

Puanlama ölçekleri, puanlamaların tutarlılığına yönelik kullanılırken öğrencileri değerlendirmek için kullanılan başarı standartlarının geçerliliklerini de geliştirir. Bu nedenle puanlama ölçekleri bir veya birden fazla ölçek kullanmalarına göre genel veya herhangi bir çalışmaya göre özel olmasına göre gruplanabilir. Analitik Puanlama ölçekleri öğrencilerin ortaya koymuş olduğu cevap veya performansına ilişkin belirli boyutların, özelliklerin veya unsurların değerlendirdirilmesini gerektirir. Bütüncül puanlama anahtarı ise bir öğrencinin ortaya koymuş olduğu cevap veya performansın niteliğinin genel olarak değerlendirilmesini gerektirir(Nikto and Brookhart, 2016: 262-264).

2.3.1.1. Analitik Dereceli Puanlama Anahtarı

Analitik dereceli puanlama anahtarı (Analitik rubrikler) iyi bir çalışmaya ait temel kriterlerin listesinin hazırlanması ve bu kriterlerin her birine uygun bir puanlama ölçeğinin hazırlanmasını gerektirir. Böylece her bir kriter için bireylere verilecek uygun puana karar verilmesi sağlanır. Genellikle öğrencilerin ortaya koyduğu çalışma puanlama dereceli puanlama anahtarında yer alan farklı derecelendirme ölçütleri ile eşleşmektedir. Bir dereceli puanlama anahtarında başarı düzeyinin puanlayıcı tarafından öğrencinin çalışmasına atanması “ en iyi yanıtı seçiniz” şeklindeki çoktan seçmeli soru gibidir. Seçilen puan öğrencinin çalışmasına en yakın açıklamaya göre belirlenmektedir. Genellikle dereceli puanlama anahtarının alt ve üst başarı düzeyleri orta düzeye göre göre daha kolay kategorileştirilebilmektedir. Bireylerin yanıtları puanlama ölçeği düzeyleriyle tutarsız bir şekilde eşleştirildiğinde puanlama sürecinin güvenilirliği azalmaktadır(Nikto and Brookhart, 2016: 264).

Analitik dereceli puanlama anahtarı öğrencinin performansının çeşitli boyutları hakkında bilgi verdiği için, öğrencilerin belirli bir alandaki zayıf ve güçlü yönleri hakkında bir görünüm sunmaktadır. Bu sayede öğrencinin yaptığı bir çalışma veya performansına yönelik ayrıntılı bir geribildirim verme olanağı sağlanmaktadır. Ayrıca analitik dereceli puanlama anahtarı ölçülecek her bir özellik için kendi ölçütleriyle bir puanlama yaptıklarından dolayı bütüncül puanlama anahtarına göre daha geçerli ve güvenilir puanlama sonuçları verebilmektedir(Kutlu vd., 2017: 60).

Bazı durumlarda analitik dereceli puanlama anahtarında bazı ölçütler diğerlerine göre daha fazla öneme sahip olabilmektedir. Bu durumda önemi fazla olan ölçütlerin altına “x 2” yazılarak öğrencinin almış olduğu puan 2 ile çarpılmaktadır. Bu sayede belirlenen ölçüt diğer ölçütlere göre daha fazla puana sahip olur(Kutlu vd., 2017: 64).

Analitik dereceli puanlama anahtarlarında her bir kriterine göre aşamalı olarak puanlar verildiğinden dolayı puanlayıcının yanlı davranması engellenebilir. Bu sayede daha nesnel ölçümler yapılabilir.

2.3.1.2. Bütünsel Puanlama Anahtarı

Bütünsel (Holistik) puanlama anahtarı belirli bir alana ait genişletilmiş yanıtlar veya önceden iyi bir çalışmanın belirlenmediği bireylerin becerilerini sentezlenmesini gerektiren çalışmalar için uygundur. Çok fazla sayıdaki çalışmanın incelenmesinde analitik dereceli puanlama anahtarı istenilen detaylı puanlamadan dolayı puanlamayı zorlaştırdığından geniş ölçek çalışmalarının değerlendirmelerinde bütünsel puanlama anahtarı daha çok tercih edilmektedir.

Bütünsel puanlama anahtarları oluşturulurken iyi bir çalışmaya yönelik puanlamada kullanılacak kriterlerin belirlenmesi önem taşımaktadır. Analitik dereceli puanlama anahtarı ile bütünsel puanlama anahtarı arasındaki fark, herhangi bir performans veya çalışma performans seviyelerindeki her bir kriter analitik dereceli puanlama anahtarı ayrı değerlendirilirken bütünsel puanlama anahtarında performansın düzeyleri bütün kriterlerle birlikte ele alınır. Bu şekilde öğrencinin ortaya koyduğu en uygun çalışma verilecek puanlamayı belirlemektedir(Nikto and Brookhart, 2016: 264).

Bütünsel puanlama anahtarı bir çalışma hakkında kısa sürede karar verme ve puanlama için daha etkindir. Ayrıca bireylerin bir konu hakkındaki genel anlamda performansına ilişkin bilgilerin elde edilmesi, bireyin o konu hakkındaki yeteneklerine ilişkin daha kapsamlı bilgi sunabilir. Ancak puanlamada kullanılan ölçütler arttıkça bütünsel puanlama anahtarlarında kullanımı zorlaşmaktadır. Aynı şekilde dikkat edilen ölçütler içerisinde, bireyin ortaya koymuş olduğu performans veya çalışmada farklılıklar artarsa bütünsel puanlama anahtarlarının kullanımı zorlaşır. Böyle durumlarda analitik puanlama anahtarlarının kullanılması daha uygundur. Çünkü kullanılan kriterlerin ağırlıklarının belirlenmesi ve ayrı ayrı puanlamanın yapılması daha kolaydır(Bahar, vd., 2015: 52-53).

Bütünsel puanlama anahtarında bireyin ortaya koymuş olduğu performans veya çalışmaya tek bir puan verilmekte ve her bir kriterde performans veya çalışmanın kalitesini

belirleyen tanımlamalar bulunmaktadır. Bu durumdaki puanlama anahtarında bireyin performans veya çalışmalarının bütününe odaklanıldığı için bazı küçük hatalar göz ardı edilebilmektedir. Bu puanlama yolunu kullanan puanlayıcı performans veya çalışmayı bir bütün olarak değerlendirip o performans veya çalışmanın hangi başarı düzeyinde olduğuna karar vermelidir(Kutlu vd., 2017: 56).

2.4. İlgili Çalışmalar

Bu bölümde Klasik Test Kuramı ve Genellenebilirlik Kuramını göre güvenilirlik belirleme yöntemlerini karşılaştıran çalışmalara yer verilmiştir.

Atılğan (2004), Müzik öğretmenliği özel yetenek seçme sınavlarında 2002 ve 2003 yıllarında toplamış olduğu verileri kullanarak G Kuramı ve Çok Değişken Kaynaklı Rasch Modeli gerçek veriler üzerinde uygulamıştır. Elde edilen verilerin analizi sonucunda G kuramının tek değişkenli ve çok değişkenli modellerinin istatistik sonuçları karşılaştırılmış ve G kuramının alternatif Karar çalışmalarının gerçek durumla tutarlılığı incelenmiştir. Bir diğer çalışma G Kuramı ve Çok Değişken Kaynaklı Rasch Modeli istatistiklerinin karşılaştırılmasıdır. Yapılan analizler sonucunda G Kuramı ve Çok Değişken Kaynaklı Rasch Modeli'nin ikisinin de kullanışlı sonuçlar ortaya koyduğu görülmektedir. G Kuramı ve Çok Değişken Kaynaklı Rasch Modeli yardımıyla değişkenlik kaynaklarının kestirilen sonuçlarının kısmen tutarlı sonuçlar verdiği gözlemlenmiştir.

Yelboğa (2007), İş performansı ölçeğini kullanarak 2005-2006 yılları arasında elde ettiği verilerle Klasik Test Kuramı ve Genellenebilirlik kuramındaki güvenilirlik kat sayılarını karşılaştırmıştır. Yapılan analizler sonucunda İş performansı ölçeğinden elde edilen sonuçların Klasik test kuramı ve Genellenebilirlik kuramının çok değişkenli modeline göre güvenilirlik katsayılarının uyumlu sonuçlar ortaya koyduğu görülmüştür.

Güler (2008), matematik başarılarının ölçülmesinden elde edilen güvenilirliklerin karşılaştırılması amacıyla KTK, G Kuramı ve Rasch Modeli üzerinden çalışma gerçekleştirmiştir. Yapılan çalışma sonucunda bu üç kuram üzerinden elde edilen puanların güvenilirlikleri hesaplanarak, güvenilirlik sonuçları karşılaştırılmıştır. KTK'da güvenilirlik için Cronbach alfa kat sayısı, puanlayıcılar arası uyumun incelenmesinde Kendall'in konkordans katsayısı, puanlayıcılar arası korelasyon katsayısı ve puanlayıcıların verdikleri puanların ortalamaları arasında farklı olup olmadığını F testi ile incelemiştir. G Kuramında b x m x p tümüyle çaprazlanmış desen kullanılarak genellenebilirlik katsayısı ve güvenilirlik katsayısı

hesaplanmıştır. Çok değişkenlik kaynaklı Rasch modelinde ise birey, puanlayıcı ve madde boyutlarına ait güvenilirlik hesaplamaları yapılmıştır. Yapılan araştırma sonucunda kullanılan ölçme aracının güvenilir sonuçlar verdiği ve puanlayıcıların puanlamada birbirleriyle uyumlu puanlar verdiği görülmüştür. Ayrıca yapılacak bir araştırmada en az iki kuramın birlikte kullanılmasının güvenilir sonuçlar elde edilmesi açısından önemi ifade edilmiştir.

Güler ve Gelbal (2010), matematik başarısını ölçme aracından toplanan puanların güvenilirliğinin belirlenmesinde KTK ve G kuramı kullanılmıştır. 2007 yılında 203 öğrenciye TIMSS-1999'un açık uçlu matematik sorularından 24 tanesinin uygulanmasıyla elde edilen puanların analizleri sonucunda, puanların iç tutarlılık değeri 0.92 bulunmuştur. Puanlayıcılar arası uyum için bakılan Kendall'in uyum katsayısı ise 0.52 şekilde elde edilmiştir. Matematik başarısının ölçülmesi amacıyla kullanılan G kuramında genellenebilirlik katsayısı 0.92 olarak bulunmuştur. Elde edilen güvenilirlik katsayısı ise 0.90 şeklindedir. Puanlayıcıların değişkenlik kaynağı, toplam varyansın %2,1'ini açıklamıştır. Bütün sonuçlar incelendiğinde matematik başarısını ölçme aracının öğrencilerin matematik başarısını belirlemede güvenilir sonuçlar verdiği görülmüştür. Puanlayıcıların puan ortalamaları arasında fark ortaya çıkmasına rağmen, puanlarının birbiriyle tutarlı olduğu sonucuna ulaşılmıştır.

Yelboğa ve Tavşancıl (2010), 2005-2006 yıllarında iş performansı ölçeğinden toplanan verilerle KTK ve G kuramı analizi gerçekleştirilmiştir. Analizler sonucunda KTK ve G kuramından kestirilen güvenilirlik katsayıları karşılaştırılmıştır. KTK'da Cronbach alfa iç tutarlılık katsayısı, test tekrar test güvenilirlik katsayısı ve puanlayıcılar arası güvenilirlik için Kendall'in uyum katsayısı; G kuramında ise çok değişkenli modelde G ve phi katsayıları kestirilmiş ve bu katsayıların birbirleri arasındaki tutarlılıkları incelenmiştir. Araştırma sonucunda KTK ve G kuramında aynı ölçme durumları için çok değişkenli modelde elde edilen güvenilirlik katsayılarının birbirleriyle uyumlu sonuçlar ortaya koyduğu görülmüştür.

Öztürk (2011), yaptığı araştırmada araştırma grubunu 2009-2010 yılları arasında Türkiye Voleybol Federasyonu tarafından daha önce belirlenmiş 10 bölgede görevli olan antrenörlerin seçtiği 102 erkek voleybolcu oluşturmaktadır. Voleybolcuların teknik becerilerini ölçmek amacıyla kullanılan ölçme aracından elde edilen puanlar KTK ve G kuramına göre incelenmiştir ve belirlenen güvenilirlik katsayıları düzeyleri birbirleriyle karşılaştırılmıştır. Elde edilen sonuçlarda güvenilirlik katsayıları iç ölçütlere göre incelendiğinde G katsayısı ve Cronbach alfa katsayısı değerinin beklenenden düşük olduğu görülmektedir. Aynı şekilde incelenen Phi güvenilirlik katsayısı ile Kendall uyum katsayılarının da beklenen değerden daha düşük değerlerde olduğu tespit edilmiştir.

Büyükkıdık (2012), yapmış olduğu araştırmada rutin olmayan (sıradışı) matematik problemlerini çözmeye becerilerinin değerlendirilmesi amacıyla hazırlanan iki performans görevinin, bütünsel ve analitik dereceli puanlama anahtarı ile 4 değerlendirici tarafından puanlanmasında elde edilen verilerin KTK ve G kuramı ile puanlayıcılar arası güvenilirliklerini incelemiştir. Araştırma da her iki kuram arasındaki avantajları ve güvenilirlik katsayıları arasındaki farklılıkları incelenmeye çalışılmıştır. KTK 'da puanlayıcılar arası güvenilirlik analizi için sınıf içi ilişki katsayısı ve puanlayıcı arası ilişki katsayısı hesaplamıştır. Aynı puanlayıcıların aynı performansa bütünsel ve analitik puanlama anahtarıyla verdikleri puan ortalamaları arasında fark olup olmadığına bağımlı gruplar t testi ile incelenmiştir. G kuramı b x g x p iki yüzeyli çaprazlanmış desen ve iki farklı puanlama anahtarının değişkenlik kaynağı olarak alındığı için b x g x p x a üç yüzeyli tümüyle çaprazlanmış desen kullanılarak güvenilirlik çalışması yapılmıştır. Araştırma sonucunda G kuramından elde edilen katsayıların KTK'da elde edilen katsayılara göre daha yüksek olduğu görülmüştür. Ayrıca G kuramının KTK'ya görece daha detaylı bilgi sunduğu ve her iki kuramda elde edilen verilerin güvenilirlik değerlerinin yüksek olduğu görülmüştür. G kuramındaki desenlerin güvenirliliği etkilediği, değişkenlik kaynağının sayısı artırıldığında bireyden kaynaklanan toplam varyansın yüzdesinin azaldığı görülmüştür.

Yıldıztekin (2014), 2013-2014 eğitim-öğretim yılı bahar döneminde Ankara ilindeki bir ortaokulda 7. Sınıfta matematik öğrenimi gören 84 öğrenciye uygulanan ve problem çözmeye becerisini ölçen açık uçlu soruların bütünsel ve analitik puanlama anahtarını kullanarak 5 puanlayıcı (matematik öğretmeni) tarafından puanlama yapılarak araştırmasını gerçekleştirmiştir. Elde edilen puanları güvenilirlik kestirimleri ve puanlayıcılar arası tutarlılık dereceleri KTK ve G kuramı ile belirlenmeye çalışılmıştır. Elde edilen veriler ile KTK'dan Pearson momentler çarpımı korelasyon katsayısı, Cronbach alfa katsayısı, Spearman sıra farkları korelasyon katsayısı, Kappa ve Krippendolf katsayısı incelenmiştir. G kuramında ise b x m x p iki yüzeyli çaprazlanmış deseninde değişkenlik kaynakları belirlenerek güvenilirlik analizleri gerçekleştirilmiştir. Yapılan araştırma sonucunda KTK ve G kuramı ile elde edilen güvenilirlik katsayılarının oldukça yüksek ve birbirine paralel olduğu görülmüştür. Kappa istatistiği katsayısının ise orta düzeyde uyumu işaret ettiği görülmektedir. G kuramında b x m x p iki yüzeyli çaprazlanmış deseninde elde edilen sonuçlara bakıldığında puanlayıcıların iki farklı dereceli puanlama anahtarının kullanması sonucunda vermiş oldukları puanlar arasında değişkenliğe etki etmedikleri görülmektedir.

Bağcı (2015), 'Matematikselsel Muhakemeyi Değerlendirme Ölçeği' ile 7. sınıftaki öğrencilerin matematikselsel muhakeme performanslarının belirlenmesi için elde edilen

puanların güvenilirliğinin incelenmesi amaçlanmıştır. ‘Matematiksel Muhakemeyi Değerlendirme Ölçeği’nin üç puanlayıcı tarafından puanlanmasıyla elde edilen ölçümlerin güvenilirliği; KTK ve G kuramının çaprazlanmış ve yuvalanmış deseniyle incelenip karşılaştırılmıştır. Her iki kuramdaki yapılan güvenilirlik analizleri sonucunda elde edilen güvenilirlik katsayıları karşılaştırılıp güvenilirlik katsayıları arasındaki farklılıkların manidarlığı test edilmiştir. Değişkenleri öğrenci (ö), soru (s) ve puanlayıcı (p) olmak üzere, G kuramının $ö \times s \times p$ çaprazlanmış deseni ve $ö \times (s : p)$ yuvalanmış deseni tasarlanmıştır. Her iki desende elde edilen Phi ve G katsayıları incelendiğinde, yuvalanmış desendeki Phi ve G katsayıları değerlerinin çaprazlanmış desenden daha yüksek olduğu görülmüştür. KTK’da güvenilirlik analizinde 3 puanlayıcı için elde edilen Cronbach alfa katsayıları ise G kuramındaki her iki desende bağıl değerlendirme için elde edilen G katsayıları ile paralel olduğu görülmüştür. G kuramı ile yapılan karar çalışmasında yüzey sayılarının bağıl ve mutlak hata varyanslarına etkisi incelenmiştir. Bu da G kuramı ile yapılan çalışmaların KTK’ya göre daha detaylı bilgi sunduğunu göstermiştir.

Pekin (2015), beş puanlayıcı tarafından “Otizm Sosyal Beceriler Profili” ölçeğinin puanlanması ile elde edilen verilerin KTK ve G kuramı aracılığıyla güvenilirlik analizlerini gerçekleştirmiştir. Ankara’da özel bir eğitim ve rehabilitasyon merkezinde eğitim alan 50 otizm tanısı konulmuş çocuk ve genç ile araştırmanın çalışma grubu oluşturulmuştur. Puanlamalar 5 puanlayıcı tarafından gerçekleştirilmiştir. Araştırmada toplanan veriler KTK için Cronbach alfa iç tutarlık katsayısı, puanlayıcılar arası güvenilirlik için Kendall’in uyum katsayısı, puanlayıcılar arası ve puanlayıcıların puan arasındaki fark olup olmamasına da ilişkili örneklerde varyans analizi ile hesaplama yapılmıştır. G kuramında ise bireylerin birlikte puanlamaları ve dönüşümlü puanlamaları iki farklı senaryo üzerinden farkı iki desen tasarlanmıştır. Birey (b), madde (m) ve puanlayıcılar (p) olmak üzere $b \times m \times p$ çaprazlanmış deseni ve $(p : b) \times m$ yuvalanmış deseni için G ve K çalışmaları yapılmış olup sonuçlar birbirleriyle karşılaştırılmıştır. Araştırmada KTK’ya göre puanlayıcıların Cronbach alfa iç tutarlık katsayı değerlerinin Yüksek olduğu ancak puanlayıcılar arası korelasyon katsayılarının ve Kendall’in uyum katsayı değerinin düşük olduğu görülmüştür. G kuramı kapsamında çaprazlanmış ve yuvalanmış desenler için ayrı ayrı hesaplanan G ve Phi katsayıları incelendiğinde bu değerlerin kabul edilebilir düzeyde olduğu görülmüştür. Her iki desende de puanlayıcı sayısının artırmanın güvenilirliği artırmada daha etkili olduğuna varılmıştır.

Şalgam (2016) tarafından 7. Sınıfta okuyan öğrencilere yönelik hazırlanan “Kısa Cevaplı Matematik Yazılı Sınavı” testinden elde edilen verilerin ölçümlerinin güvenilirliğinin

incelenmesi amacıyla çalışma gerçekleştirilmiştir. Bağımsız üç puanlayıcı tarafından yazılı testlerinin puanlanmasıyla elde edilen ölçümler KTK'nın test tekrar test yöntemi ve G kuramının $b \times m \times p$ çaprazlanmış deseni kullanılarak karşılaştırılmıştır. İki hafta arayla öğrencilere uygulanan Kısa Cevaplı Matematik Yazılı Sınavı aynı puanlayıcılar tarafından puanlanmıştır. Yapılan araştırma neticesinde her iki kuramdan elde edilen güvenilirlik katsayıları karşılaştırıldığında aralarında benzer sonuçlar olduğu gözlemlenmiştir.

Kızıltoprak (2016) tarafından yapılan araştırmanın amacı PISA matematik okuryazarlığı maddelerinin yeterliliklerini belirlemek amacıyla geliştirilmiş olan "Yeterlilik Şeması" adlı puanlama aracıyla elde edilen puanların ölçüt geçerliği, yapı geçerliği ve güvenilirliğinin belirlenmesidir. Bu nedenle ilk olarak Yeterlik Şeması puanlama anahtarı'nın Türkçe'ye uyarlama çalışması gerçekleştirilmiştir. Türkçe'ye uyarlanan Yeterlik Şeması puanlama anahtarı kullanılarak yayınlanmış bütün PISA matematik okuryazarlığı maddeleri 5 puanlayıcıdan tarafından bir ay arayla iki kez puanlanmıştır. Elde edilen verilerden puanlamaların güvenilirliğini belirlemek amacıyla KTK'da puanlayıcılar arası ve puanlayıcılar içi güvenilirlik katsayıları incelenmiştir. G kuramında ise tümüyle çaprazlanmış desen tek değişkenli ve çok değişkenli modellerde G ve K çalışmaları incelenmiştir. Yapılan analizler sonucunda Yeterlik Şeması puanlama anahtarından elde edilen puanlamaların maddelerin deneysel güçlük puanlarındaki gözlenen değişkenliğin %74,7'si oranında anlamlı olarak açıkladığı görülmüştür. KTK'da puanlayıcı içi güvenilirlik için hesaplanan korelasyon katsayılarının kategoriler ve puanlayıcılar bazında farklı sonuçlar verdiği, puanlayıcılar arası güvenilirlikte ise güvenilirlik değerinin güçlü olduğu görülmüştür. G kuramındaki çok değişkenli model üzerinden yapılan G ve K çalışmalarıyla kestirilen G ve Phi katsayılarının tek değişkenli model ile kestirilen G ve Phi değerlerinden yüksek olduğu görülmüştür.

Yılmaz (2019) tarafından yapılan çalışmada Zarar Verici Sosyal Davranışlar ölçeği kullanılarak puanlayıcılar arası güvenilirliğin KTK ve G Kuramına göre karşılaştırılıp hangi kuramın detaylı bilgi sunduğunun belirlenmesi amaçlanmıştır. Kullanılan ölçekten elde edilen verilerin uyum değerleri KTK çerçevesinde Cronbach alfa iç tutarlılık katsayısı, Pearson momentler katsayısı ve Kendall'in uyum katsayısı ile hesaplanmıştır. Puanlayıcıların vermiş olduğu puanlar arasındaki ilişkinin belirlenmesi amacıyla ilişkili örneklemlerde t-testi analizi yapılmıştır. G kuramında ise $b \times m \times p$ çaprazlanmış desen ile $(p : b) \times m$ yuvalanmış desenler tasarlanmış ve G ile K çalışmaları yapılmıştır. Yapılan çalışmada beklenildiği gibi G kuramının KTK'ya göre daha detaylı bilgi sunduğu görülmüştür.

KTK ve G kuramına göre yapılan araştırmalar incelendiğinde, G kuramının KTK'ya göre daha ayrıntılı bilgi verdiği görülmüştür ve güvenilirlik katsayılarının KTK'ya göre daha

yüksek deęer aldıęı sonucuna ulaşılmıřtır. Ayrıca deęiřkenlerin artırılması durumunda G kuramına göre ölçme objesinden kaynaklanan varyansın azaldıęı ifade edilmektedir.



ÜÇÜNCÜ BÖLÜM

YÖNTEM

3.1. Araştırma Modeli

Bu çalışmada Ata model uçak puanlama anahtarı kullanılarak güvenilirliği belirleme de Genellenebilirlik kuramının tümüyle çaprazlanmış (u x m x p) deseni ile güvenilirlik kestirimleri incelenmeye çalışılacaktır. Bu çerçevede çalışmanın betimsel araştırma türünde olduğu söylenebilir. Betimsel araştırmalar, verilen bir durumun dikkatli ve eksiksiz bir şekilde tanımlanması sağlar. Eğitim alanındaki en uygun araştırma modeli olan betimsel yöntem tarama çalışması, birey, grup veya fiziki ortamlara ait özellikleri özetler. Betimsel araştırmalar bir öğrenci grubunun başarısını belirlemek, bir grubun davranışlarını veya tutumlarını tanımlamak için kullanılabilir(Büyüköztürk, Kılıç Çakmak, Akgün, Karadeniz ve Demirel, 2018: 24).

3.2. Çalışma Grubu

Çalışma grubunun belirlenmesinde seçkisiz olmayan örnekleme yöntemlerinden ölçüt örnekleme yöntemi kullanılmıştır. Ölçüt örnekleme yönteminde araştırma gözlem birimlerinin belirli nitelikteki kişi, nesne, olay veya durumlardan meydana gelebilir. Böyle bir durumda örnekleme için belirlenen ölçütleri karşılayan birimler örnekleme dâhil edilirler(Büyüköztürk vd., 2018: 95).

Araştırmada uçaklar için belirlenen ölçüt bütün uçakların kaplamalarının tamamlanarak birleştirilmesi ve uçuşa hazır olmasıdır. Puanlayıcılar için ise temel kriter puanlamaya hakim bir yeterliliğe sahip olmasıdır. Bu amaçla daha önce birey veya durumlar üzerinde puanlama yapılmaları yapılacak puanlamada kolaylık sağlayacaktır.

Araştırma grubunu oluşturan maket uçakların tamamı 2020-2021 yılları arasında Gaziantep'teki farklı okullarda yapılmıştır. Maket uçakları yapan bireylerin tamamı 10-14 yaşları arasında olup 32 saatlik model uçak eğitimi alarak puanlanacak uçaklarını tamamlamıştır. Model uçakları puanlamak içinde 5 öğretmen gönüllü olup puanlamayı tamamlamıştır. Puanlayıcılara ilgili bilgiler Tablo 3'te verilmiştir.

Tablo 3. Puanlayıcılara ait bilgiler

Puanla yıcı	İkamet Ettikleri İl	İş Tecrübesi	Görevi
1	Şırnak/İdil	5 yıl	Sosyal Bilgiler Öğretmeni
2	Şırnak/İdil	5 yıl	Fizik Öğretmeni
3	Şırnak/İdil	6 yıl	Antrenör
4	Şırnak/İdil	6 yıl	Beden Eğitimi Öğretmeni
5	Şırnak/İdil	10 yıl	Beden Eğitimi Öğretmeni

Tablo 3 incelendiğinde puanlayıcıların kendi alanlarında en az 5 yıl tecrübeli oldukları görülmektedir.

Puanlayıcılar model uçak dereceli puanlama anahtarı ile puanlamaya geçmeden önce dereceli puanlama anahtarı üzerindeki her bir kriter için nasıl puanlaması gerektiği hakkında sunum yapılmıştır. Daha sonra hazırlanan bir ata model uçak üzerinde puanlamalar yapılarak ölçülecek kriter ile dereceli puanlama anahtarı üzerindeki puanlanacak kısımlar arasındaki ilişki detaylıca gösterilmiştir. Yaklaşık bir yapılan saatlik sunum sonucunda puanlayıcıların hazır olmasıyla puanlama sürecine başlanmıştır.

3.3. Veri Toplama Aracı

Araştırmada kullanılmaya uygun olarak bir analitik puanlama anahtarı geliştirilmiştir. Bu çalışmada katılımcı olan 5 puanlayıcının seçiminde gönüllülük esas alınmıştır. Model uçak için ata model analitik puanlama anahtarını kullanacak 5 puanlayıcıya puanlamanın nasıl yapılacağı hakkında gerekli eğitim verilmiştir.

3.3.1. Model Uçak Dereceli Puanlama Anahtarı Geliştirilme Süreci

Araştırma verilerinin elde edilmesi amacıyla hazırlanan analitik puanlama anahtarı geliştirme sürecinde HBÖGM tarafından hazırlanan Model Uçak (Ata) Çerçeve Kurs

Programı incelenmiştir. Analitik dereceli puanlama anahtarının geliştirme aşamasında takip edilmesi gereken adımları şu şekilde ifade etmiştir (Kutlu vd. 2017: 69-70);

1. Analitik dereceli puanlama anahtarı hazırlarken bireyin ortaya koyacağı performans alt boyutlarına ayrılmalıdır. Bu alt boyutlar belirlenirken ortaya konulan performans ve ölçülmek istenen zihinsel süreçlere dikkat edilmelidir. Bireyin veya ürünün hangi özellikleri gözlemlenmek veya geliştirilmek isteniyorsa ona göre alt boyutların belirlenmesi gerekmektedir.

2. Belirlenen alt boyutlarda bireyin gösterebileceği başarı düzeyleri belirlenerek bu düzeyler bir puan ile nitelendirilir. Başarı düzeyleri ölçülecek özelliğin hassaslığına göre artırılıp azaltılabilir. Analitik dereceli puanlama anahtarlarında başarı düzeyleri arttıkça ölçülecek özellik hakkında daha hassas ölçümler yapılabilir. Ancak diğer taraftan bu dereceli puanlama anahtarının kullanılabilirliğini olumsuz yönde etkileyebilir. Çünkü çok fazla başarı düzeyinin olması hem hazırlama aşamasında hem de kullanılmasında zaman alıcı ve zordur. Bu sebeple 3 ile 5 arasında başarı düzeyinde olan analitik puanlama anahtarları daha çok tercih edilmektedir.

3. Ölçülecek özelliğin alt boyutundaki her düzey için ayrıntılı tanımlamalar yapılır. Tanımlamaların anlaşılır olması puanlama güvenilirliğini ve doğruluğunu artırır. Bu nedenle bu tanımlamalar yapılırken göreceği ifadelerden özellikle kaçınılmalıdır. Çünkü bu tür ifadeler her puanlayıcı tarafından farklı anlaşılacağından ölçme aracının güvenilirliğini azaltacaktır.

Bu ifadelerden yola çıkarak aşağıdaki adımlar takip edilerek araştırmada kullanılacak analitik dereceli puanlama anahtarı elde edilmiştir.

3.3.1.1. Ölçütlerin Belirlenmesi

Analitik dereceli puanlama anahtarındaki ölçütler belirlenirken Model Uçak (Ata) Çerçeve Kurs Programı'ndaki ölçütler dikkate alınmıştır. Bir model uçak uçuşa hazır hale gelip iyi uçuşması için sırasıyla;

1. Orta kanat genişliğinin projede belirtildiği gibi 64.5 cm olması gerekmektedir. Kanat genişliğinin bu ölçülerden çok veya az olması uçuş performansını etkileyecektir.

2. Plan üzerinde üç kısım olarak inşa edilen kanat ucunun plandaki 6 cm yükseklik ölçüsüne göre destek çıtalılarıyla desteklenerek kanat büküntüsünü verecek şekilde kanat uçlarının orta kanada birleştirilmesi gerekmektedir.

3. Gövde ıtası plastik buruna takılarak plana uygun olarak 54,5 cm olarak kesilmelidir.

4. Pellur kağıdı ile kaplanacak olan orta kanat ve dümenlerinin kaplama esnasında yırtılmamasına özen gösterilmelidir. Aksi halde boşluklara sızan hava tanecikleri uçuşu olumsuz etkileyecektir.

5. Kaplama tamamlandıktan sonra daha önce hazırlanan sprey su ile orta kanat ve dümen kaplamalarının ıslatılması ve bir zemine sabitlenmesi gerekmektedir. Bu sayede kaplama gerilerek büzölmelerin önüne geçilmiş olacak ve havada süzölürken iyi bir performans sağlayacaktır. Kaplanan alanların büzölmüş olması girintilere sebep olacağından uçuş performansı ve yüzeydeki rüzgâr akışını etkileyecektir.

6. Dönüş dümeninin yükseliş dümeninin üzerine 90° lik bir açı ile yapıştırılması gerekir. Böylece uçağın yön değıştirmeden daha uzun süre uçuşması sağlanabilir.

7. Ata model uçağın plandaki ağırlık merkezinin (Center Gravity) CG kontrol edilmesi, modele önden bakıldığında kanat ve yükseliş dümeninin birbirine paralelliğinin kontrol edilmesi, model uçağın kanat ve gövdesi yere paralel bir şekilde uçuşmasının kontrolü kriterlerinin dikkate alınması doğru bir uçuşun gerçekleşmesini sağlayacaktır.

8. Uçuş denemelerinde uçağın kaplamasının yere inişte yırtılması veya iskelet ıtarlarının yırtılması yapılan bir uçağın dayanıklılığı ve hazırlama aşamasında uygun yapıştırma aşamalarının yapıp yapılmadığı hakkında bize bir bilgi sunacaktır.

9. Uçuş süresine bakıldığında daha önce açık alanda yapılan denemelerde rüzgâr etkisinden dolayı 10 saniye ve üzerinde uçuşların gerçekleştiğı ölçölmüştür. Ancak rüzgâr hızı her an aynı hızda esmediğinden bu durumdaki bir hatayı engellemek adına kapalı spor salonunda uçuşlar gerçekleştirilmiştir.

3.3.1.2. Maddeler İçin Uzman Görüşüne Başvurulması

Analitik dereceli puanlama anahtarının hazırlanması aşamasında model uçak alanında en az 3 yıl eğitim tecrübesi bulunan 2 model uçak eğitmeni, dil bilgisi alanında uzman 1 Türkçe öğretmeni, ölçme değıerlendirme alanında uzman 1 araştırma görevlisi ve 1 doçent olmak üzere 5 uzman ile çalışılmıştır.

3.3.1.3. Deneme Formunun Hazırlanması

Model Uçak (Ata) Çerçeve Kurs Programı ve uzmanların vermiş olduğu yönergeler dâhilinde analitik dereceli puanlama anahtarı taslak formunun maddeleri oluşturulmuştur. Hazırlanan dereceli puanlama anahtarı 10 maddeden oluşmaktadır ve her madde 4 dört farklı düzeyde puanlanacak şekilde hazırlanmıştır. Maddelerden en düşük 1, en yüksek 4 olmak üzere dereceli puanlama anahtarından toplamda 40 puan alınabilir.

3.3.1.4. Deneme Formunun Uygulanması

Deneme uygulaması 5 puanlayıcı tarafından her bir uçak ayrı ayrı puanlanacak şekilde 16 uçak üzerinde yapılmıştır. Ölçülecek uçakların uçuş süresinin rüzgârın esme hızından etkilenmemesi için kapalı spor salonunda çalışmalar gerçekleştirilmiştir. 16 uçağın 10 madde ile 5 puanlayıcı tarafından puanlanması sonucunda elde edilen sonuçlar incelendiğinde G katsayısının .21 ve Phi katsayısının .10 olduğu görülmüştür. Deneme çalışmasında uçakların iskelet kısımlarının kırılmaması ve kaplamaların yırtılmaması sebebiyle Puanlayıcıların birbirine benzer puanlar vermesi 8. ve 9 maddeler analizden çıkarılmış ve bu maddeler analizden çıkarıldığında G katsayısının .29 ve Phi katsayısının .11 olduğu görülmüştür. Madde değişkeninden kaynaklanan varyansın %47.6 ile en yüksek varyansa sahip olması sebebiyle maddelerde yer alan uzaklık birimleri azaltılarak daha hassas ölçümler yapılması hedeflenmiştir.

3.3.1.5. Nihai Formların Oluşturulması Uygulamanın Yapılması

Puanlayıcıların vermiş olduğu yanıtlar incelendiğinde puanlayıcıların verdiği puanların hazırlanan Ata Model Uçak Puanlama Anahtarı ölçütlerindeki düzeylerinde yığılmalar gösterdiği fark edilmiştir. Buradan yola çıkarak daha önce hazırlanan Ata Model Uçak Puanlama Anahtarı'ndaki yığılmalar gösteren ölçütlerinde bulunan düzeylerindeki değerler daha hassas ölçümler yapılması amacıyla değerler arasındaki uzunluk azaltılmıştır. Nihayetinde yapılan değişiklikler sonucunda 8 ölçüt ve her ölçütün 4 düzeyden oluştuğu Ata Model Uçak Puanlama Anahtarı elde edilmiştir.

Tablo 4. Ata Model Uçak Puanlama Anahtarı

ÖLÇÜTLER	ZAYIF	ORTA	İYİ	ÇOK İYİ	PUAN
Orta kanat genişliği	+1 veya -1 cm'den fazla sapma gösterdi. ()	+1 veya -1 cm sapma gösterdi. ()	+5 veya -5 cm sapma gösterdi. ()	64-65 cm aralığında ayarlandı. ()	
Kanat ucu yüksekliği	+1 veya -1 cm'den fazla sapma gösterdi. ()	+1 veya -1 cm sapma gösterdi. ()	+0.5 veya -0.5 cm sapma gösterdi. ()	6 cm olarak ayarlandı. ()	
Model uçak gövde uzunluğu	+1 veya -1 cm'den fazla sapma gösterdi. ()	+1 veya -1 cm sapma gösterdi. ()	+5 veya -5 cm sapma gösterdi. ()	54-55 cm uzunlukta olacak şekilde ayarlandı. ()	
Kaplama yapım aşamasında	Üç veya daha fazla yerden yırtıldı. ()	İki yerden yırtıldı. ()	Bir yerden yırtıldı. ()	Hiç yırtılma görülmedi. ()	
Kaplamanın düzgünlüğü	Üç veya daha fazla yerden büzülme oldu. ()	2 yerden büzülme oldu. ()	Bir yerde büzülme oldu. ()	Hiçbir büzülme olmadı. ()	
Dönüş ve yükseliş dümeni yapıştırılma açısı	87 derece açıdan daha az açı ile yapıştırıldı. ()	87-88 derece açı ile yapıştırıldı. ()	88-89 derece açı ile yapıştırıldı. ()	89-90 derece açı ile yapıştırıldı. ()	
Uçağı uçuşa hazırlama aşamasında denge ayarları	Uçak için herhangi bir denge sağlanamamıştır. ()	Uçak sağa veya sola doğru yatmaktadır. ()	Uçak öne veya arkada doğru yatmaktadır. ()	Uçak tam dengededir. ()	
Uçuş süresi	0-2 saniye uçuş gerçekleştirdi. ()	2-4 saniye uçuş gerçekleştirdi. ()	4-6 saniye uçuş gerçekleştirdi. ()	6 saniye ve üstü süre uçuş gerçekleştirdi. ()	

Tablo 4 'teki 8 kriterlerin her bir düzeyine ait özellikler belirlenerek dereceli puanlama anahtarı son halini almıştır.

3.4. Verilerin Analizi

Ata model uçaklarına ait Ata Model Puanlama Anahtarının değerlendirmesini yapabilecek 5 puanlayıcı verilen eğitim sonucunda puanlama görevinde bulunmuştur. Bu puanlayıcılar birbirinden tamamen bağımsız puanlama yapma amacıyla her bir puanlayıcıya ayrı ayrı aynı marka kalem, dümen açılarını ölçmek için açıölçer ve uzunlukları ölçmek için metre verilmiştir. Ayrıca puanlayıcıların birbirinden etkilenmemeleri için aralarında belirli mesafenin bulunmasına özen gösterilmiştir. Verilerin analizi iki ayrı aşamada gerçekleşmiştir. Birinci aşamada 120 uçağın 5 puanlayıcıya göre 8 madde verilen yanıtlara göre betimsel istatistikleri hesaplanmıştır. İkinci aşamada ise elde edilen veriler araştırmanın amacına göre genellenebilirlik kuramı için tümüyle çaprazlanmış (u x m x p) desene uygun olarak analizleri yapılmıştır. Verilerin analizine başlamadan önce 5 puanlayıcının yapmış olduğu puanlama sonucunda elde edilen veriler incelendiğinde betimsel istatistikler **Tablo 5** 'te verilmiştir.

Tablo 5. 5 Puanlayıcıdan Elde Edilen Puanlara Ait Betimsel İstatistikler

İstatistikler	Puanlayıcılar				
	1	2	3	4	5
n_u	120	120	120	120	120
n_m	8	8	8	8	8
Minimum	16	16	14	15	17
Maksimum	28	29	30	28	29
Ortalama	22.73	23.2	22.1	23.6	22.6
Std. Sapma	2.59	2.57	3.06	2.77	2.76
Çarpıklık	-0.092	-	-	-	-
Basıklık	-0.386	0.278	0.208	0.549	0.205
		0.224	0.235	0.111	0.561

Tablo 5'e bakıldığında en yüksek ortalamanın 4. puanlayıcıya (23.63), ve en düşük ortalamanın 2. puanlayıcıya ait olduğu görülmekle birlikte bütün ortalama puanların birbirine yakın değerler aldığı görülmektedir. Bütün puanlayıcıların yapmış olduğu puanlamalar sonucu elde edilen çarpıklık ve basıklık değerleri incelendiğinde her iki katsayının negatif değerler aldığı görülmekle birlikte -1 ile +1 arasında değer aldığı görülmektedir. Elde edilen değerlerin negatif olması puanlayıcıların çarpıklık değerinin sola çarpık olduğu şeklinde yorumlanmaktadır. Basıklık değerinin negatif olması ise puanlayıcıların vermiş olduğu puanların normalden daha basık bir dağılıma sahip olduğunu gösterir. Çarpıklık ve basıklık değerlerinin -1 ile +1 arasında değer alması ise puanlayıcılara ait puanların normal dağıldığına kanıt olarak gösterilebilir. Araştırmaya ait betimsel analizler SPSS 22 istatistik programı ve Genellenabilirlik kuramına ait analizler ise EDUG 6.1-e istatistik programı ile yapılmıştır.

DÖRDÜNCÜ BÖLÜM

BULGULAR VE YORUM

Bu bölümde, araştırma sürecinde belirlenen alt problemlere ait bulgular verilmiş ve elde edilen bulgulara dayalı yorumlamalar yapılmıştır.

4.1. Birinci Alt Probleme Ait Elde Edilen Bulgular ve Yorumlar

G çalışması sonucunda kestirilen varyans bileşenleri ve toplam varyansı açıklama yüzdeleri nasıl değişmektedir?

Ata model uçak dereceli puanlama anahtarı kullanılarak 120 Ata model uçak, 8 kriter doğrultusunda 5 puanlayıcı tarafından puanlanmıştır. Belirlenen desen (uçak, madde ve puanlayıcı) tamamen çaprazlanmış desen, her puanlayıcının her bir uçağı aynı değerlendirme formunu kullanmasıyla oluşturulmuş olan $u \times m \times p$ deseni (u: uçak, m: madde, p: puanlayıcı) ile G çalışması yapılmıştır. G çalışmasında $u \times m \times p$ deseni sonucunda kestirilen varyans bileşenleri ve toplam varyans yüzdeleri Tablo 4'te verilmiştir.

Tablo 6. $u \times m \times p$ Çaprazlanmış Desenine Ait G Çalışması Sonucunda Kestirilen Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri

Varyans Kaynağı	Kareler toplamı (S.S.)	sd	Kareler	Varyans	%
			ortalaması (M.S.)		
u	325.829	119	2.738	0.017	1.6
p	18.844	4	4.711	0	0.0
m	1140.363	7	162.909	0.260	24.3
up	239.306	476	0.503	0.008	0.8
um	1652.688	833	1.984	0.309	28.9
pm	141.990	28	5.071	0.039	3.6
ump,e	1457.460	3332	0.437	0.437	40.8
Toplam	4976.479	4799			100%

Tablo 6 'da verilen $u \times m \times p$ çaprazlanmış desenine ait G çalışması sonucunda kestirilen varyans bileşenleri ve toplam varyansı açıklama yüzdeleri incelendiğinde, uçak (u) ana etkisine ait varyans bileşeninin toplam varyansın % 1.6'sını açıkladığı görülmektedir. Uçaklara ait varyans bileşenleri uçakların özellikleri hakkında bilgiler sunar. **Tablo 6** 'da görüldüğü gibi uçaklara ait varyans bileşeni ($\sigma^2(u) = 0.017$) oldukça küçük bir değer çıkmıştır.

Uçaklara ait varyans bileşeninin toplam varyansı açıklama oranına göre küçük olması, ölçülen uçaklarının birbirinden farklılaşmadığını göstermekte, uçakların ölçülen özellikleri bakımından benzer olduğu söylenebilir.

Puanlayıcı ana etkisi puanlama esnasında bazı puanlayıcıların birey veya ürünü değerlendirmede diğer puanlayıcılara göre katı veya cömert olup olmaması konusunda bilgi verir(Shavelson and Webb, 1991). **Tablo 6** incelendiğinde puanlayıcı (p) ana etkisi için kestirilen varyans bileşeni toplam varyansın %0.0'ını açıkladığı görülmektedir. Puanlayıcı için kestirilen varyans bileşenin sıfır veya sıfıra yakın olması, puanlayıcıların yapmış olduğu puanlamalar arasında herhangi bir farklılaşmanın bulunmadığını, aynı şekilde puanlayıcılar vermiş olduğu puanlar arasında bir tutarlılığın olduğu şeklinde ifade edilebilir.

Madde ana etkisi için kestirilen varyans bileşeni incelendiğinde toplam varyansın %24.3'ünü açıklamaktadır. Madde ana etkisine ait varyans bileşeni toplam varyans içinde en fazla üçüncü varyans değerine ($\sigma^2(g)= 0.260$) sahip olduğu görülmektedir. Madde ana etkisine ait varyans bileşenlerinin yüksek olması madde güçlükleri arasında farklılığın olduğunu göstermektedir. Bir testteki madde veya görevlerin madde güçlüklerinin farklı olması belirlenen madde örnekleminin evrene genellenmesinde oldukça önemlidir.

Uçak x puanlayıcı (up) ortak etkileşiminde ortaya çıkan varyans bileşeni toplam varyansın %0.8'ini açıklamaktadır. Uçak x puanlayıcı ortak etkileşimi ile kestirilen varyans bileşeni puanlayıcıların bireyi veya ürünü puanlamamadaki tutarlı olup olmamasını ortaya koyar(Shavelson and Webb, 1991). Araştırmada kestirilen uçak x puanlayıcı ortak etkisine ait varyansın küçük olması, uçakların değerlendirilmesinde puanlayıcıdan puanlayıcıya pek farklılık göstermediğini ifade etmektedir.

En yüksek ikinci varyans yüzdesi uçak x madde (um) ortak etkileşimindeki değişkenlik kaynağına ait olduğu görülmektedir. Uçak x madde (um) ortak etkileşimindeki ortaya çıkan varyans toplam varyansın %28.9'unu açıklamaktadır. Uçak x madde etkileşimi uçaklarının ölçülen özelliklerinin maddelere göre farklılık gösterip göstermediğini ifade eder (Shavelson and Webb, 1991). **Tablo 6** incelendiğinde uçak x madde etkileşiminde ortaya çıkan varyansın yüksek olması bazı uçakların ölçülen özelliklerinin maddeden maddeye farklılık gösterdiğini göstermektedir.

Puanlayıcı x madde ortak etkisinden ortaya çıkan varyans incelendiğinde toplam varyansın % 3.6'sını açıkladığı görülmektedir. Puanlayıcı x madde (pm) ortak etkileşimi puanlayıcının

maddeden maddeye kararlılıkları hakkında bilgi verir (Shavelson and Webb, 1991). Puanlayıcı x madde (pm) ortak etkileşiminde ortaya çıkan varyans oranının düşük olması maddelere verilen puanların puanlayıcıdan puanlayıcıya çok fazla farklılaşmadığını göstermektedir. Bu durum puanlayıcıların hazırlanan model uçakları puanlarken maddeden maddeye kararlı puanladıkları şeklinde yorumlanabilir.

Uçak x madde x puanlayıcı (upm) ortak etkisinden ortaya çıkan varyans bileşeni artık varyans olarak ifade edilir. **Tablo 6** incelendiğinde artık varyans bileşeni toplam varyansın % 40.8'ini açıklaması bakımından değişkenler arasında en büyüğüdür. Yapılan çalışmada artık varyans bileşenini büyük olması Uçak x madde x puanlayıcı ortak etkileşimi veya tesadüfi hata kaynaklarının büyük olabileceği şeklinde yorumlanabilir.

4.2. İkinci Alt Probleme Ait Elde Edilen Bulgular ve Yorumlar

G çalışması sonucunda kestirilen G ve Phi katsayı değerleri nedir?

Tablo 7. u x m x p tümüyle çaprazlanmış desenine göre G ve Phi katsayıları

Katsayılar	Değer
G katsayısı	0.25
Phi katsayısı	0.17

Tablo 7 'de u x m x p tümüyle çaprazlanmış desenine göre G ve Phi katsayıları incelendiğinde, bağıl hata varyansına ($\sigma^2(\delta)= 0.05123$) göre G katsayısının .25, ölçme objesinin tek başına ortaya koyduğu varyansın dışında kalan bütün varyans bileşenlerinin toplamı olarak hesaplanan mutlak varyansına ($\sigma^2(\Delta)= 0.08476$) göre Phi katsayısının .17 olduğu görülmektedir.

4.3. Üçüncü Alt Probleme Ait Elde Edilen Bulgular ve Yorumlar

Puanlayıcı sayılarının ve madde sayılarının değişmesi durumlarında; K çalışmasında kestirilen G katsayısı ile bağıl ve mutlak hata varyansları nasıl değişmektedir?

Öğrenciler tarafından yapılan 120 adet uçak belirlenen 8 madde ile 5 puanlayıcı tarafından değerlendirilmiştir. Bütün değişkenlerin çaprazlandığı u x m x p deseninde uçaklar görev nesnesi olarak belirlenip puanlayıcı sayısının sabit tutulup madde sayısının artırılması, madde sayısının sabit tutulup puanlayıcı sayısının artırılması ve puanlayıcı sayısı ile madde sayısının her ikisinin de artırılmasıyla yapılar senaryolar için kestirilen genellenebilirlik katsayısı (G), bağıl hata varyansı ($\sigma^2(\delta)$) ve mutlak hata varyansına ($\sigma^2(\Delta)$) ait değerler **Tablo 8'** de verilmiştir.

Tablo 8. u x m x p Tümüyle Çaprazlanmış Desenine Göre Puanlayıcı veya Madde Sayılarının Artırılmasıyla Yapılan K Çalışması Sonucu Elde Edilen Bulgular

u x m x p				
n_p	n_m	G	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
5	8	0.252	0.051	0.085
5	10	0.294	0.041	0.068
5	15	0.380	0.028	0.046
5	20	0.445	0.021	0.035
5	25	0.496	0.018	0.028
5	30	0.537	0.015	0.024
6	8	0.259	0.049	0.083
7	8	0.265	0.048	0.081
8	8	0.270	0.047	0.080
9	8	0.274	0.046	0.079
10	8	0.277	0.045	0.078
6	12	0.341	0.033	0.055
7	14	0.383	0.028	0.047
8	16	0.420	0.024	0.040
9	18	0.453	0.021	0.036
10	20	0.482	0.019	0.032

n_p : puanlayıcı sayısı n_m : madde sayısı

Yapılan araştırmada 120 uçağın 8 madde doğrultusunda 5 puanlayıcı tarafından birbirinden bağımsız olarak puanlanmasına göre yapılan karar çalışmasında G katsayısı 0.252 olarak hesaplanmıştır. **Tablo 8'**deki verilere göre puanlayıcı sayısı ($n_p = 5$) sabit tutulup madde sayısı ($n_m = 10$) iki artırıldığında G katsayısının 0.294 olduğu görülmektedir. Aynı şekilde puanlayıcı sayısının ($n_p = 5$) sabit tutulup madde sayısının sırasıyla 15, 20, 25 ve 30'a çıkarılmasıyla elde edilen G katsayısının değerleri sırasıyla 0.380, 0.445, 0.496 ve 0.537

olduğu görülmektedir. Puanlayıcı sayısının ($n_p = 5$) sabit tutulup madde sayısının artırılması sonucu oluşturulan senaryolara göre kestirilen G katsayısı değerlerinin arttığı görülmektedir. Elde edilen G katsayıları arasındaki artışın fazla olması madde sayılarının artırmanın G katsayısını artıracığı şeklinde yorumlanabilir.

Tablo 8 incelendiğinde madde sayısının ($n_m = 8$) sabit tutulup puanlayıcı sayısının 6 olması durumunda G katsayısının 0.260 olduğu görülmektedir. Puanlayıcı sayısının bir artırılması durumunda G katsayısının araştırmada kullanılan asıl verilere göre 0.008 şeklinde arttığı görülmektedir. Madde sayısının ($n_m = 8$) sabit tutulup puanlayıcı sayısının sırasıyla 7,8,9 ve 10 olması durumunda ise G katsayısının alacağı değerler sırasıyla; 0.265, 0.270, 0.274 ve 0.277 olduğu görülmektedir. Madde sayısının sabit tutularak puanlayıcı sayısının birer artırılması durumunda G katsayısının artırılmasında önemli bir artış sağlanmadığı görülmüştür.

Araştırmada elde edilen veriler üzerinde yapılan senaryolara göre puanlayıcı sayısının 6 ve madde sayısının 12 olması durumunda G katsayısının 0.341 olduğu görülmektedir ve puanlayıcı sayısı ile madde sayısındaki artışın G katsayısında 0.090 kadar artış sağladığı görülmektedir. Yine puanlayıcı sayısının sırasıyla 7, 8, 9 ve 10 olması durumunda G katsayısında 0.383, 0.420, 0.453 ve 0.482 şeklinde artış olduğu görülmektedir. Elde edilen veriler doğrultusunda madde ve puanlayıcı sayısının birlikte artırılması durumunda G katsayısında artış sağladığı görülmektedir.

Tablo 8'de puanlayıcı sayısının sabit tutulup, madde sayısının artırılması durumunda bağıl hata varyansı ($\sigma^2(\delta)$) ve mutlak hata varyansı ($\sigma^2(\Delta)$) incelendiğinde, madde sayılarındaki artışa bağlı olarak bağıl hata ve mutlak hata varyanslarında azalma olduğu görülmektedir. Örneğin $n_p = 5$ ve $n_m = 8$ olduğu durumda bağıl hata varyansı 0.051 ve mutlak hata varyansı 0.085 olarak kestirilmişken, $n_p = 5$ ve $n_m = 30$ olduğu durumda ise bağıl hata varyansı 0.015 ve mutlak hata varyansı 0.024 olarak hesaplanmıştır. Madde sayısının sabit tutulup puanlayıcı sayısının artırılması durumunda ise bağıl hata ve mutlak hata varyanslarında önemli derecede düşüş olmadığı görülmüştür. **Tablo 8**'deki senaryo incelendiğinde $n_p = 6$ ve $n_m = 8$ olduğu durumda bağıl hata varyansı 0.049 ve mutlak hata varyansı 0.083 olduğu görülmüştür. $n_p = 10$ ve $n_m = 8$ olduğu durumda ise bağıl hata varyansının 0.045 ve mutlak hata varyansının 0.078 değeri aldığı görülmektedir.

BEŞİNCİ BÖLÜM

SONUÇ VE ÖNERİLER

5.1. Sonuçlar

Birinci alt probleme ait sonuçlar: *G çalışması sonucunda kestirilen varyans bileşenleri ve toplam varyansı açıklama yüzdeleri nasıl değişmektedir?*

Uçak x madde x puanlayıcı (ump) ortak etkisi veya artık etkisinin (e) iki yüzeyli çaprazlanmış desen ile G kuramına göre yapılan analize bakıldığında toplam varyans açıklama yüzleri arasında en büyük etkiye sahip olduğu sonucuna varılmıştır. Elde edilen bu sonuç, uçak, madde ve puanlayıcı ortak etkisinin büyük bir boyutta olduğunu veya ölçme işleminde kontrol edilemeyen tesadüfi hataların yüksek oranda olduğunu göstermektedir.

Uçak x madde (um) değişkenlik kaynağının ortak etkisi incelendiğinde en yüksek ikinci varyans değerine sahip olduğu görülmektedir. İncelenen varyans değerinin yüksek olması ölçülen uçakların ölçme sonuçlarının birbirleriyle karşılaştırıldığında maddeden maddeye göre oldukça değiştiğini göstermektedir. Bu sonuçtan elde edilen bulgulara göre puanlama esnasında bir maddenin kolay değerinin zor geldiği sonucuna varılabilir.

Kullanılan desenlerde en yüksek üçüncü varyans değerine sahip olan madde (m) değişkenlik kaynağıdır. Bulgulara göre varyans yüzde değerinin yüksek çıkması madde güçlüklerinin farklılaştığı sonucuna varılabilir. Madde güçlülüklerinin farklı olması madde örnekleminin madde evrenine genellenmesinde önemli bir yer tutmaktadır.

u x m x p tümüyle çaprazlanmış desende puanlayıcı x madde (pm) ortak etkisi incelendiğinde puanlayıcı x madde ortak etkisine ait varyans yüzdesinin düşük olduğu görülmektedir. Değişkenlik kaynağına ait varyans yüzdesi değerinin düşük olması, puanlayıcıların maddeleri değerlendirirken benzer şekilde cevaplandığı diğer bir ifadeyle maddeleri tutarlı bir şekilde puanladıkları sonucuna varılabilir.

Değişkenlik kaynakları arasında bulunan uçak ana etkisine ait elde edilen bulgular incelendiğinde, ölçme objesi olarak seçilen uçakların varyans açıklama yüzdesinin düşük olduğu görülmüştür. Böyle bir durum öğrenciler tarafından yapılan uçakların birbirine benzer özellikte yapıldığı şeklinde yorumlanabilir.

Uçak x puanlayıcı (up) ortak etkisine ait varyans değerinin düşük olduğu görülmüştür. Bu durum uçakların ölçme sonuçlarının puanlayıcılar arasında farklılık yaratmadığı sonucunu göstermektedir.

Son olarak puanlayıcı (p) deęişkenlik kaynağına yönelik incelenen bulgular sonucunda puanlayıcı deęişkenlik kaynağına ait varyans yüzdesinin sıfır olduęu görülmüştür. Bu durum beklenen bir durum olmak birlikte puanlayıcılar arasında tutarlılıęın yüksek olduęunu gösterir.

Öztürk'ün (2011) yapmış olduęu araştırmada birey ana etkisine ait varyansın düşük (%2.5.) olduęu görülmüştür. Birey ana etkisine ait varyansın düşük olması bireyler arasındaki farklılıkların ortaya çıkmadığına göstergesidir. Yapılan çalışmada benzer şekilde uçak ana etkisinden ortaya çıkan varyansın düşük olması uçakların birbirine oldukça benzer şekilde yapıldığını göstermektedir. Aynı şekilde görev temel etkisinden kaynaklanan düşük (%0.5) olması görevlerin güçlük düzeylerinin birbirine oldukça yakın olduęunun göstergesidir. Araştırmada elde edilen bulgulara bakıldığında madde temel etkisinden kaynaklanan varyans oranının yüksek (%24.3) olduęu görülmektedir. Bu sonuçlar maddelerin güçlük düzeylerinin birbirinden farklılık gösterdiğini ifade etmektedir.

İkinci alt probleme ait sonuçlar: *G çalışması sonucunda kestirilen G ve Phi katsayı değerleri nedir?*

Tümüyle çaprazlanmış u x m x p deseni üzerinden uçakların ölçme objesi olarak seçilerek yapılan G çalışması sonucunda G katsayısı 0.25 ve Phi katsayısı 0.17 olarak elde edilmiştir. Sonuçlar incelendiğinde Phi güvenilirlik katsayısının ve G (genellenebilirlik) katsayısının istenilen düzeyden daha düşük düzeyde olduęu görülmüştür. Maddelerden kaynaklanan varyansın toplam varyans oranına göre yüksek olması nedeniyle madde sayısının artırılması genellenebilirlik veya güvenilirlik katsayılarını artırabilmektedir.

Uygulanan testlerin bireyler (uçak) veya puanlayıcılar arasında farklılıkları ortaya koymaması G katsayısının düşük değer almasına ve maddelerin uçakları ölçmedięi sonucunu ortaya çıkarmaktadır. Demirbilek Zorba'nın (2020) yapmış olduęu çalışma bu araştırmayı destekler niteliktedir.

Üçüncü alt probleme ait sonuçlar: *Puanlayıcı sayılarının ve madde sayılarının deęişmesi durumlarında; K çalışmasında kestirilen G katsayısı ile baęlı ve mutlak hata varyansları nasıl deęişmektedir?*

Model uçakların değerlendirilmesinde 120 uçağın 8 madde ve bu maddelerin 5 puanlayıcı tarafından puanlanmasıyla elde edilen G katsayısı 0.254, baęlı hata varyansı ($\sigma^2(\delta)= 0.051$) ve mutlak varyansı ($\sigma^2(\Delta)= 0.085$) şeklinde bulunmuştur. **Tablo 8**

incelendiğinde madde sayısının ($n_m = 8$) sabit tutulup puanlayıcı sayısının artırılması durumunda G katsayısında veya bağıl ve mutlak hata varyanslarında ciddi bir artışın olmadığı görülmektedir. Bu durumun sebebi puanlayıcıdan kaynaklanan varyans değerinin sıfır olmasından kaynaklanmaktadır. Puanlayıcıdan kaynaklanan varyansın düşük olması puanlayıcı sayısındaki artış ve azalışta G katsayısını artma ve azalmasının daha az etkileyecektir. Puanlayıcı sayısının sabit tutulup ($n_p = 5$) madde sayısındaki artış ise G katsayısının artmasını ve bağıl ile mutlak hata varyanslarında ise azalma meydana getirmektedir. Maddelerden kaynaklanan varyansın puanlayıcıdan kaynaklanan varyanstan yüksek olması sebebiyle maddelerdeki artış veya azalış G katsayısını, bağıl ve mutlak hata varyanslarını daha fazla etkileyecektir. Madde sayısının puanlayıcı sayısının iki katı olması durumunda ise aynı durumda G katsayısında artış sağladığı, bağıl ve mutlak hata varyanslarında ise belirli düzeyde azalış sağladığı görülmektedir. Ancak karar çalışmalarında yapılan senaryolar incelendiğinde G katsayısının istenilen düzeyde olmadığı sonucuna ulaşılmıştır.

Güneş'in (2020) yapmış olduğu çalışmada puanlayıcı sayısının 5 madde sayısının 8 olduğu durumda $G = 0.79$, madde sayısının 24'e çıkarılması durumunda $G = 0.80$ olduğu ve madde sayısının artırılması durumunda G katsayısında yüksek bir fark olmadığı görülmektedir. Yapılan çalışma sonucunda ise puanlayıcı sayısının 5 olduğu durumda madde sayısının artırılması durumunda G katsayısının önemli ölçüde arttığı sonucuna varılmıştır.

Öztürk'ün (2011) araştırmasında yapmış olduğu karar çalışmasında puanlayıcı ve madde sayısının artırılması durumunda G katsayısında da önemli ölçüde artış olduğu görülmektedir. Yapılan bu karar çalışmadaki artış miktarı araştırmamızı destekler niteliktedir.

5.2. Öneriler

1. Ata model uçakların değerlendirmesinde puanlayıcılar arası güvenilirlikte Genellebilirlik katsayısının düşük olduğundan madde ve puanlayıcı sayısının artırılması istenilen düzeyde güvenilir ölçmeler yapmak için uygun olabilir.

2. Puanlama yapacak puanlayıcıların daha önce model uçak eğitimi almaları daha hassas ölçmeler yapması açısından olumlu katkı sağlayabilir.

3. Ata model uçakların deęerlendirmesinde puanlayıcıların, puanlamaya yönelik açık, anlaşılır ve yeterli düzeyde bilgilendirilmesi veya uygulamalarla eğitim verilmesi puanlayıcılar arası uyumu artırabilir.

4. Madde örnekleme seçilirken madde evrenini temsil eden maddelerin seçilmesi ölçülecek özelliklerin daha doğru bir şekilde ölçülmesine yardımcı olur. Yapılan çalışmadan yola çıkarak madde sayısının artırılması ile ölçülecek özelliklerin çoğalması maddelerin madde evrenini temsil etmesi açısından uygun olabilir.

5. Yapılan bu çalışma $u \times m \times p$ çaprazlanmış deseni göre analiz edilmiştir. Yapılacak benzer çalışma kuramın farklı deęişkenlik kaynakları veya modelleri ile ele alınarak yapılabilir.

6. Yeni Araştırmalarda yapılan uçaklarla birlikte uçakların yapım sürecine yönelik puanlamalar yapmak için maddelerin eklenmesiyle geliştirilecek dereceli puanlama anahtarının kullanılması uygun olabilir.

KAYNAKLAR

- Arıkan, S., Çelen, Ü., Gülleroğlu, H. D., Gültekin, S., Kilmen, S. ve Köse, İ. A. (2014). *Eğitimde Ölçme ve Değerlendirme* (3. baskı). Ankara: Edge Akademi
- Atılğan H. (2004). *Genellenebilirlik Kuramı Ve Çok Değişkenlik Kaynaklı Rasch Modelinin Karşılaştırılmasına İlişkin Bir Araştırma*. Doktora Tezi, Ankara: Hacettepe Üniversitesi.
- Atılğan, H. (2019). *Genellenebilirlik Kuramı ve Uygulaması* (1. baskı). Ankara: Anı Yayıncılık.
- Bağcı, V. (2015). *Matematiksel Muhakeme Becerisinin Ölçülmesinde Klasik Test Kuramı ile Genellenebilirlik Kuramındaki Farklı Desenlerin Karşılaştırılması*. Yayınlanmamış yüksek lisans tezi. Ankara: Gazi Üniversitesi.
- Bahar, M., Nartgün, Z., Durmuş, S. ve Bıçak, B. (2015). *Geleneksel-Tamamlayıcı Ölçme ve Değerlendirme Teknikleri Öğretmen El Kitabı* (7. baskı). Ankara: Pegem Akademi.
- Baker, F. B. (2016). *Madde Tepki Kuramının Temelleri*. (N. Güler, Dü., ve M. İlhan, Çev.) Ankara: Pegem Akademi.
- Baykul, Y. (2015). *Eğitimde ve Psikolojide Ölçme: Klasik Test Teorisi ve Uygulaması* (3. baskı). Ankara: Pegem Akademi.
- Brennan, R. L. (2001). *Generalizability Theory*. Springer-Verlag New York, Inc.
- Brennan, R. L. (2011). *Generalizability theory and classical test theory*. Applied Measurement And Education,1-21
- Büyükkıdık, S. (2012). *Problem Çözme Becerisinin Değerlendirilmesinde Puanlayıcılar Arası Güvenirliğin Klasik Test Kuramı Ve Genellenebilirlik Kuramına Göre Karşılaştırılması*. Yayınlanmamış yüksek lisans tezi. Ankara: Hacettepe Üniversitesi.
- Büyüköztürk Ş., Kılıç Çakmak, E., Akgün, Ö. E., Karadeniz, Ş. ve Demirel, F. (2018). *Eğitimde Bilimsel Araştırma Yöntemleri* (25. baskı). Ankara: Pegem Akademi.
- Büyüköztürk Ş., Çokluk, Ö. ve Köklü, N., (2016). *Sosyal Bilimler İçin İstatistik* (18. baskı). Ankara: Pegem Akademi.
- Büyüköztürk, Ş., Kılıç Çakmak, E., Akgün, Ö. E., Karadeniz, Ş. ve Demirel, F. (2018). *Eğitimde Bilimsel Araştırma Yöntemleri* (25. baskı). Ankara: Pegem Akademi.

- Can, A. (2018). *SPSS ile Bilimsel Araştırma Sürecinde Nicel Veri Analizi*. (6. baskı). Ankara: Pegem Akademi.
- Cardinet, J., Johnson, S., and Pini, G. (2010). *Applying generalizability theory using EduG*. USA: Routledge-Taylor and Francis Group.
- Crocker, L. and Algina, J. (2006). *Introduction to classical and modern test theory*. USA: Cengage Learning
- DeMars, C. (2010). *Item Response Theory*. USA: Oxford University Press.
- DeMars, C. (2016). *Madde Tepki Kuramı*. E. H. Özberk, ve H. Kelecioğlu (Dü.) içinde, Bölüm 1. Giriş (s. 1-30). Ankara: Nobel Akademi.
- Erkuş, A., Sünbül, Ö., Sünbül, S.Ö., Yormaz, S. ve Aşiret S., (2017). *Psikolojide Ölçme ve Ölçek Geliştirme-II* (1. baskı). Ankara: Pegem Akademi.
- Gulliksen, H. (1950). *Theory of Mental Tests*. New York: John Willey and Sons, Inc
- Güler, N. (2008). *Klasik Test Kuramı Genellenabilirlik Kuramı ve Rasch Modeli Üzerine Bir Araştırma*. Yayınlanmamış doktora tezi. Ankara: Hacettepe Üniversitesi
- Güler N. (2019). *Eğitimde Ölçme ve Değerlendirme*. (13. Baskı). Ankara: Pegem Akademi.
- Güler, N., Kaya Uyanık, G. ve Taşdelen Teker, G. (2012). *Genellenebilirlik Kuramı* (1. baskı). Ankara: Pegem Akademi.
- Hambleton, R. K., and Jones, R. W. (1993). *Comparison of classical test theory and item response theory and their applications to test development*. Educational Measurement(12), 38-47.
- Hambleton, R. K., Swaminathan, H. (1985). *Item Response Theory Principles and Applications*. Netherlands: Springer Netherlands
- Hambleton, R. K., Swaminathan, H., and Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. USA: Sage Publications.
- HBÖGM, (2017). *Model Uçak (Ata) Çerçeve Kurs Programı*. Ankara: MEB
- Karagöz, Y. (2017). *SPSS ve AMOS Uygulamalı Nitel- Nicel- Karma Bilimsel Araştırma Yöntemleri ve Yayın Etiği*. Ankara: Nobel Yayın Dağıtım.

- Kızıltoprak, F. (2016). *Pısa Matematik Okuryazarlığı Maddelerine İlişkin Yeterlilik Şemasıyla Elde Edilen Ölçümlerin Güvenirliğinin Genellenebilirlik ve Klasik Test Kuramına Dayalı Olarak Karşılaştırılması* Ankara: Gazi Üniversitesi.
- Kutlu, Ö., Doğan, C., ve Karakaya, İ. (2017). *Ölçme ve Değerlendirme Performansa ve Portfolyoya Dayalı Durum Belirleme*. 5. Baskı. Ankara: Pegem Akademi.
- Lester, P. E., Inman, D. and Bishop, L. K. (2014) *Handbook of Tests and Measurement In Education and The Social Sciences*. United States of America: Rowman and Littlefield Publishers.
- Lord, F. (1980). *Applications of item response theory to practical testing problems*. New Jersey: Lawrence Erlbaum Associates Publishers.
- Lord, F. M., and Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading MA: Addison- Wesley.
- Nitko, J.A and Brookhart S.M. (2016). *Educational Assessment of Student*. San Antonio: Pearson
- Özçelik, D.A. (2013). *Okullarda Ölçme ve Değerlendirme Öğretmen El Kitabı* (2. Basım). Ankara: Pegem Akademi.
- Öztürk, M. E. (2011). *Voleybol Becerileri Gözlem Formu İle Elde Edilen Puanların Genellenebilirlik Ve Klasik Test Kuramına Göre Karşılaştırılması*. Yayınlanmamış yüksek lisans tezi. Ankara: Hacettepe Üniversitesi.
- Pekin, Z. (2015). *Otizm Sosyal Beceriler Profili Ölçeğinde Puanlayıcılar Arası Güvenirliğin Klasik Test Kuramı ve Genellenebilirlik Kuramına Göre Karşılaştırılması* Yayınlanmamış yüksek lisans tezi. Ankara: Hacettepe Üniversitesi.
- Shavelson, R. J. and Webb, N. M. (1991). *Generalizability Theory: A Primer*. Sage Publications, USA
- Spearman, C. (1927). *The Abilities of Man Their Nature and Measurement*. New York: The MacMillan Company
- Şalgam, A. (2016). *Kısa Cevaplı Matematik Yazılı Sınavının genellenebilirlik Kuramı ve Test Tekrar Test yöntemiyle Güvenirliğinin Kıyaslanması* Ankara: Gazi Üniversitesi.
- Tavşancıl, E. (2002). *Tutumların ölçülmesi ve SPSS ile veri analizi*. Ankara: Nobel.

- Tekindal, S. (2017). *Okullarda Ölçme ve Değerlendirme Yöntemleri* (6. basım). Ankara: Nobel Akademik Basım Yayıncılık Dağıtım.
- THK (2017). *Türk Kuşu Kursiyer El Kitabı*. Ankara: Model Uçak Okulu Müdürlüğü
- Thorndike, R. (2014) *Measurement and Evaluation in Psychology and Education-* Harlow: Pearson Education Limited
- TTKB (2005) *İlköğretim 1-5. Sınıf programları tanıtım el kitabı* , Ankara: TC MEB Talim ve Terbiye Kurulu Başkanlığı .
- Yelboğa, A. (2007). *Klasik Test Kuramı ve Genellenebilirlik Kuramına Göre Güvenirliğin Bir İş Performansı Ölçeği Üzerinde İncelenmesi*. Doktora Tezi, Ankara: Ankara Üniversitesi.
- Yelboğa, A., ve Tavşancıl, E. (2010). *Klasik Test ve Genellenebilirlik Kuramına Göre Güvenirliğin Bir İş Performansı Ölçeği Üzerinde İncelenmesi*. Kuram ve Uygulamada Eğitim Bilimleri, 10(3), 1825-1854.
- Yıldıztekin, B. (2014), *Klasik Test Kuramı ve Genellenebilirlik Kuramına Göre Puanlayıcılar Arası Tutarlığın Karşılaştırılması*. Yayınlanmamış yüksek lisans tezi. Ankara: Hacettepe Üniversitesi.
- Yılmaz, K. (2019). *Zarar Verici Sosyal Davranışlar Ölçeğinin Puanlayıcılar Arası Güvenirliğinin Klasik Test Kuramı ve Genellenebilirlik Kuramına Göre Karşılaştırılması* Yayınlanmamış yüksek lisans tezi. Bolu: Bolu Abant İzzet Baysal Üniversitesi

EKLER

EK-1. Ata model uçak puanlama anahtarı

Uçağın adı-soyadı:

Puanlayıcının adı-soyadı:

ATA MODEL UÇAK PUANLAMA ANAHTARI

ÖLÇÜTLER	ZAYIF	ORTA	İYİ	ÇOK İYİ	PUAN
Orta kanat genişliği	+1 veya -1 cm'den fazla sapma gösterdi. ()	+1 veya -1 cm sapma gösterdi. ()	+5 veya -5 cm sapma gösterdi. ()	64-65 cm aralığında ayarlandı. ()	
Kanat ucu yüksekliği	+1 veya -1 cm'den fazla sapma gösterdi. ()	+1 veya -1 cm sapma gösterdi. ()	+0.5 veya -0.5 cm sapma gösterdi. ()	6 cm olarak ayarlandı. ()	
Model uçak gövde uzunluğu	+1 veya -1 cm'den fazla sapma gösterdi. ()	+1 veya -1 cm sapma gösterdi. ()	+5 veya -5 cm sapma gösterdi. ()	54-55 cm uzunlukta olacak şekilde ayarlandı. ()	
Kaplama yapım aşamasında	Üç veya daha fazla yerden yarıldı. ()	İki yerden yarıldı. ()	Bir yerden yarıldı. ()	Hiç yırtılma görülmedi. ()	
Kaplamanın düzgünlüğü	Üç veya daha fazla yerden büzülme oldu. ()	2 yerden büzülme oldu. ()	Bir yerde büzülme oldu. ()	Hiçbir büzülme olmadı. ()	
Dönüş ve yükseliş dümeni yapıştırılma açısı	87 derece açıdan daha az açı ile yapıştırıldı. ()	87-88 derece açı ile yapıştırıldı. ()	88-89 derece açı ile yapıştırıldı. ()	89-90 derece açı ile yapıştırıldı. ()	
Uçağı uçuşa hazırlama aşamasında denge ayarları	Uçak için herhangi bir denge sağlanamamıştır. ()	Uçak sağa veya sola doğru yatmaktadır. ()	Uçak öne veya arkada doğru yatmaktadır. ()	Uçak tam dengededir. ()	
Uçuş süresi	0-2 saniye uçuş gerçekleştirildi. ()	2-4 saniye uçuş gerçekleştirildi. ()	4-6 saniye uçuş gerçekleştirildi. ()	6 saniye ve üstü süre uçuş gerçekleştirildi. ()	