

**OTİZM SOSYAL BECERİLER PROFİLİ ÖLÇEĞİNDE  
PUANLAYICILAR ARASI GÜVENİRLİĞİN KLASİK TEST  
KURAMI VE GENELLENEBİLİRLİK KURAMINA GÖRE  
KARŞILAŞTIRILMASI**

**COMPARISON OF INTERRATER RELIABILITY BASED ON  
CLASSICAL TEST THEORY AND GENERALIZABILITY  
THEORY FOR AUTISM SOCIAL SKILLS PROFILE**

**Zeynep PEKİN**

Hacettepe Üniversitesi

Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin

Eğitim Bilimleri Anabilim Dalı, Eğitimde Ölçme ve Değerlendirme Bilim Dalı İçin  
Öngördüğü

Yüksek Lisans Tezi

olarak hazırlanmıştır.

2015

Eđitim Bilimleri Enstitüsü M¼d¼rl¼đ¼'ne,

Zeynep PEKİN'in hazırladıđı "Otizm Sosyal Beceriler Profili ¼lçeđinde Puanlayıcılar Arası Güvenirliđin Klasik Test Kuramı ve Genellenebilirlik Kuramına G¼re Karşılařtırılması" bařlıklı bu alıřma j¼rimiz tarafından **Eđitim Bilimleri Anabilim Dalı, Eđitimde ¼lme ve Deđerlendirme Bilim Dalı'nda Y¼ksek Lisans** olarak kabul edilmiřtir.

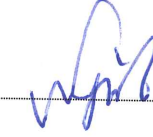
Bařkan Prof. Dr. Selahattin GELBAL



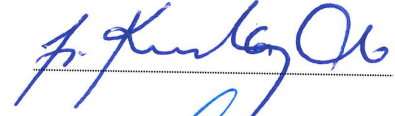
¼ye (Danıřman) Yrd. Do. Dr. Sevda ETİN



¼ye Do. Dr. Neře G¼LER



¼ye Do. Dr. İsmail KARAKAYA



¼ye Yrd. Do. Dr. Derya OBANOĐLU AKTAN



## ONAY

Bu tez Hacettepe ¼niversitesi Lisans¼st¼ Eđitim-¼đretim ve Sınav Y¼netmeliđi'nin ilgili maddeleri uyarınca yukarıdaki j¼ri ¼yeleri tarafından ..... / ..... / ..... tarihinde uygun g¼r¼lm¼ř ve Enstit¼ Y¼netim Kurulunca ..... / ..... / ..... tarihinde kabul edilmiřtir.

Prof. Dr. Berrin AKMAN  
Eđitim Bilimleri Enstit¼s¼ M¼d¼r¼

# OTİZM SOSYAL BECERİLER PROFİLİ ÖLÇEĞİNDE PUANLAYICILAR ARASI GÜVENİRLİĞİN KLASİK TEST KURAMI VE GENELLENEBİLİRLİK KURAMINA GÖRE KARŞILAŞTIRILMASI

**Zeynep PEKİN**

**ÖZ**

Bu araştırmada, “Otizm Sosyal Beceriler Profili” (OSBP) ölçeğinin 5 puanlayıcı tarafından puanlanması ile elde edilen puanlara klasik test kuramı ve genellenebilirlik kuramı ile puanlayıcılar arası güvenirlik karşılaştırılması yapılmıştır. Genellenebilirlik kuramında puanlayıcıların birlikte ve dönüşümlü puanlama yapmasıyla oluşturulan farklı desenlerden ve klasik test kuramından elde edilen güvenirlik katsayılarının düzeyleri saptanmış ve hangi kuramın daha fazla bilgi sunduğu belirlenmeye çalışılmıştır.

Araştırmanın çalışma grubunu Ankara ilindeki bir özel eğitim ve rehabilitasyon merkezinde eğitim almakta olan 50 otizmlı çocuk ve genç oluşturmaktadır. Ayrıca, araştırmada 50 otizmlı çocuk ve genci Otizm Sosyal Beceriler Profili ölçeğinin “sosyal karşılıklık” alt boyutunda yer alan 15 madde doğrultusunda puanlayan 5 puanlayıcı yer almıştır. Elde edilen veriler klasik test kuramında puanlayıcıların iç tutarlılık güvenirliği Cronbach Alfa ( $\alpha$ ) katsayısı, puanlayıcılar arası güvenirlik Kendall’ın uyum katsayısı, puanlayıcılar arası korelasyon katsayısı ve puanlayıcıların verdikleri puanlar arasında fark olup olmadığı ise ilişkili örneklerde varyans analizi ile hesaplanmıştır. Genellenebilirlik kuramı kapsamında ise puanlayıcıların bireyleri birlikte ve dönüşümlü puanladıkları iki farklı senaryo üzerinden iki desen tasarlanmıştır. Bireylerin (b) aynı maddeler (m) doğrultusunda puanlayıcıların (p) her biri tarafından puanlandığı  $b \times m \times p$  çapraz deseni ve bireylerin tüm maddeler doğrultusunda farklı puanlayıcılar tarafından puanlandığı  $(p : b) \times m$  yuvalanmış deseni için ayrı ayrı G ve K çalışmaları yapılmış ve sonuçlar birbirleriyle karşılaştırılmıştır.

Araştırma sonucunda klasik test kuramına göre puanlayıcıların iç tutarlılık güvenirliğini ifade eden Cronbach Alfa katsayılarının yüksek değerler aldığı, ancak puanlayıcılar arası korelasyon katsayılarının ve Kendall’ın uyum katsayısının

düşük değerler aldığı görülmüştür. Genellenebilirlik kuramı kapsamında çapraz ve yuvalanmış desen için ayrı ayrı hesaplanan G ve Phi katsayılarının kabul edilebilir düzeylerde değerler aldığı, birey ve puanlayıcı ortak etkilerine ait varyans bileşenlerinin her iki desende de toplam varyansı açıklama oranlarının yüksek olduğu görülmüştür. Her iki desende yapılan karar çalışmaları sonucunda güvenilirliği artırmak için puanlayıcı sayısını artırmanın, madde sayısını artırmaya göre daha etkili olduğu sonucuna varılmıştır.

Elde edilen tüm bu sonuçlara göre, genellenebilirlik kuramının klasik test kuramına göre daha detaylı bilgi verdiği, her iki kuram doğrultusunda Otizm Sosyal Beceriler Profili ölçeğinin otizmlili çocuk ve gençlerin sosyal beceri seviyelerini belirlemede güvenilir sonuçlar verdiği ve her iki kuram kapsamında da puanlayıcılar arası tutarlılık düzeyinin düşük olduğu belirlenmiştir.

**Anahtar sözcükler:** Klasik test kuramı, genellenebilirlik kuramı, puanlayıcılar arası güvenilirlik, Kendall'ın uyum katsayısı, sosyal becerilerin değerlendirilmesi

**Danışman:** Yrd. Doç. Dr. Sevda ÇETİN, Hacettepe Üniversitesi, Eğitim Bilimleri Anabilim Dalı, Eğitimde Ölçme ve Değerlendirme Bilim Dalı

# COMPARISON OF INTERRATER RELIABILITY BASED ON CLASSICAL TEST THEORY AND GENERALIZABILITY THEORY FOR AUTISM SOCIAL SKILLS PROFILE

**Zeynep PEKIN**

## ABSTRACT

In this study, interrater reliability was compared based on classical test theory and generalizability theory according to the scores which were obtained from five raters' ratings with Autism Social Skills Profile. Levels of reliability coefficients obtained from classical test theory and different designs in generalizability theory formed by five raters' jointly and alternatively ratings was determined and which theory presented more information was tried to be specified.

The research group consists of 50 children and youth with autism who are getting training in a special education and rehabilitation center in Ankara and 5 raters rated them through social reciprocity sub-scale under Autism Social Skills Profile (ASSP). In the classical test theory, Cronbach Alpha ( $\alpha$ ) coefficient for internal consistency, Kendall's concordance coefficient for interrater reliability and correlation coefficients of five raters' scores were calculated and it was investigated whether there was a difference among the means of raters' scores with F test. In the generalizability theory, two different designs were formed according to raters' jointly and alternatively ratings. Several G and D studies were made for crossed design  $p \times i \times r$  (p: person, i: item and r: rater) which people were scored by all raters through all items and nested design  $(r : p) \times i$  (p: person, i: item and r: rater) which people were scored by different raters through all items and the results were compared to each other.

According to the all obtained scores, in the classical test theory it was seen that Cronbach Alpha ( $\alpha$ ) coefficients were high values, on the other hand interrater correlation coefficients and Kendall's concordance coefficient were low values. In the generalizability theory, it was seen that, G and Phi coefficients for crossed and nested designs were acceptable values and person-rater common influence variance component had high percentage of the description of total variance in both designs. As a result of D studies in both designs, it was decided that

increasing the number of raters had a higher influence on increasing the reliability than the number of items.

Consequently, it was obtained that generalizability theory provided more information than classical test theory, ASSP provided reliable results in specifying social skill levels of children and youth with autism and interrater reliability level was low according to classical test theory and generalizability theory.

**Keywords:** classical test theory, generalizability theory, interrater reliability, Kendall's concordance coefficient, evaluation of social skills

**Advisor:** Yard. Doç. Dr. Sevda ÇETİN, Hacettepe University, Department of Educational Sciences, Division of Educational Measurement and Assessment

## ETİK BEYANNAMESİ

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmasında,

- tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- başkalarının eserlerinden yararlanması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi,
- kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- ve bu tezin herhangi bir bölümünü bu üniversitede veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

İmza  
Zeynep PEKİN

## TEŞEKKÜR

Tez sürecimin her aşamasında başaracağıma olan inancını açık gönüllülükle ifade eden, panik olduğum her anda tatlı diliyle beni sakinleştiren, olur olmaz saatlerde telefonlarıma cevap veren, yapıcı yorum ve önerileriyle beni yönlendiren danışmanım Yard. Doç. Dr. Sevda ÇETİN'e;

Çalışma sürecimde yardım ve desteğini esirgemeyen, sadece bir telefon uzaklığında olduğunu hissettiren hocam ve jüri üyem Doç. Dr. Neşe Güler'e;

Eleştiri ve önerileriyle çalışmama katkıda bulunan, akademisyen olarak örnek aldığım çok değerli hocalarım ve jüri üyelerim Prof. Dr. Selahattin GELBAL'a, Doç. Dr. İsmail KARAKAYA'ya, ve Yard. Doç. Dr. Derya ÇOBANOĞLU AKTAN'a;

Verilerimi toplamamda bana yardımcı olan, özel eğitimle ilgili sorularına bıkmadan usanmadan cevap veren, hayata karşı duruşuyla örnek aldığım kuzenim Damla ORUÇ BAŞESGİOĞLU'na;

Hayatımda iz bırakanlar listesinde kısa sürede üst sıralarda yer almayı başaran, stresimin çeneme vurduğu zamanlarda beni sabırla (!) dinleyen Dr. Zuhal YILMAZ'a;

Bu dönemdeki asosyalliğimi (!) anlayışla karşılayan, telefonda saatlerce stresimi paylaştığım, bana sonuna kadar inanan, pozitif enerjileriyle her an yanımda olan dostlarıma;

Hayatımın her anında koşulsuz sevgi ve destekleriyle beni şımartan, ne kadar şanslı olduğumu her gün tekrar tekrar bana hissettiren vazgeçilmezlerim Annem, Babam ve Kardeşim'e;

ÇOK TEŞEKKÜR EDERİM...



# İÇİNDEKİLER

ÖZ.....	iii
ABSTRACT.....	v
ETİK BEYANNAMESİ.....	vii
TEŞEKKÜR.....	viii
İÇİNDEKİLER.....	ix
TABLolar DİZİNİ.....	xii
ŞEKİLLER DİZİNİ.....	xiii
SİMGELER VE KISALTMALAR DİZİNİ.....	xiii
1. GİRİŞ.....	1
1.1. Sosyal Beceriler.....	1
1.2. Sosyal Becerilerin Değerlendirilmesi.....	5
1.2.1. Görüşme.....	5
1.2.2. Öz-değerlendirme.....	6
1.2.3. Sosyometri.....	6
1.2.4. Gözlem.....	7
1.2.5. Dereceleme Ölçekleri.....	9
1.2.5.1. Sayısal Derecelendirme Ölçeği.....	11
1.2.5.2. Grafikselsel Derecelendirme Ölçeği.....	11
1.2.5.2.1. Yatay Grafikselsel Derecelendirme Ölçeği.....	11
1.2.5.2.2. Dikey Grafikselsel Derecelendirme Ölçeği.....	12
1.2.5.3. Betimselsel Derecelendirme Ölçeği.....	13
1.2.5.1. Karşılaştırmalı Derecelendirme Ölçeği.....	14
1.3. Problem Durumu.....	14
1.4. Araştırmanın Amacı Ve Önemi.....	15
1.5. Problem Cümlesi.....	17
1.5.1. Alt Problemler.....	17
1.6 Sayıtlılar.....	18
1.7. Sınırlılıklar.....	18
1.8. Araştırmanın Kuramsal Temeli.....	18
1.8.1. Klasik Test Kuramı.....	18
1.8.1.1. Puanlayıcılar-arası Güvenirlik.....	20
1.8.2. Genellenebilirlik Kuramı.....	24
1.8.2.1. Genellenebilirlik (G) Çalışması & Karar (K) Çalışması.....	28
1.8.2.2. Tesadüfi Değişkenlik Kaynakları & Sabit Değişkenlik kaynakları.....	29
1.8.2.3. Çaprazlanmış (Crossed) Desen & Yuvalanmış (Nested) Desen.....	30
1.8.2.4. Bağlı (Görelî) Hata Varyansı & Mutlak Hata Varyansı.....	32
1.8.2.5. Genellenebilirlik Katsayısı (G-katsayısı) & Genellenebilirlik İndeksi (Phi-katsayısı).....	34
2. İLGİLİ ARAŞTIRMALAR.....	37
2.1. İlgili Araştırmalar Özet.....	46
3. YÖNTEM.....	47
3.1. Araştırmanın Türü.....	47

3.2. Araştırma Grubu .....	47
3.3. Veri Toplama Aracı .....	48
3.3.1. Otizm Sosyal Beceriler Profili (OSBP) .....	48
3.3.2. OSBP Türkçe Formu (OSBP-T) .....	48
3.4. Verilerin Analizi .....	50
4. BULGULAR VE TARTIŞMA .....	52
4.1. Betimsel İstatistikler .....	52
4.2. Alt Problem 1: Otizm Sosyal Beceriler Profilinden Elde Edilen Puanların Klasik Test Kuramına Göre İncelenmesi .....	53
4.2.1. Klasik Test Kuramına Göre Otizm Sosyal Beceriler Profilinden Elde Edilen Puanların 5 Puanlayıcıya İlişkin İç Tutarlılık Düzeyleri .....	53
4.2.2. Klasik Test Kuramına Göre Otizm Sosyal Beceriler Profiline İlişkin Değerlendirmede Puanlayıcılar Arası Tutarlılık Derecesi .....	53
4.3. Alt problem 2: Genellenebilirlik Kuramına Göre; Birey (b), Madde (m) Ve Puanlayıcı (p) Değişkenlerinin Çapraz Tasarlandığı $b \times m \times p$ Deseninin Sonuçları.....	55
4.3.1. $b \times m \times p$ Deseninin G Çalışması Sonucunda Kestirilen Varyans Bileşenleri Ve Toplam Varyansı Açıklama Yüzdeleri.....	55
4.3.2. $b \times m \times p$ Deseninde Puanlayıcı ve Madde Sayılarının Artırılıp Azaltılmasıyla Yapılan Karar Çalışması Sonuçları .....	57
4.4. Alt problem 3: Genellenebilirlik Kuramına Göre; Birey (b) Ve Puanlayıcı (p) Değişkenlerinin Yuvalandığı, Madde (m) Değişkeninin ise Çapraz Tasarlandığı $(p : b) \times m$ Deseninin Sonuçları.....	59
4.4.1. $(p : b) \times m$ Deseninin G Çalışması Sonucunda Kestirilen Varyans Bileşenleri Ve Toplam Varyansı Açıklama Yüzdeleri.....	59
4.4.2. $(p : b) \times m$ Deseninde Puanlayıcı ve Madde Sayılarının Artırılıp Azaltılmasıyla Yapılan Karar Çalışmaları Sonuçları .....	61
4.5. Alt Problem 4: $b \times m \times p$ ve $(p : b) \times m$ Desenlerinden Elde Edilen G Çalışması Sonuçlarının Karşılaştırılması .....	62
4.6. Alt Problem 5: $b \times m \times p$ ve $(p : b) \times m$ Desenlerinde Puanlayıcı ve Madde Sayılarının Artırılıp Azaltılmasıyla Yapılan Karar Çalışmaları Sonuçlarının Karşılaştırılması .....	64
5.SONUÇ ve ÖNERİLER .....	68
5.1. Sonuçlar.....	68
5.1.1. Birinci Alt Probleme İlişkin Sonuçlar .....	68
5.1.2. İkinci Alt Probleme İlişkin Sonuçlar.....	68
5.1.3. Üçüncü Alt Probleme İlişkin Sonuçlar.....	69
5.1.4. Dördüncü Alt Probleme İlişkin Sonuçlar .....	71
5.1.5. Beşinci Alt Probleme İlişkin Sonuçlar .....	72
5.2. Öneriler.....	72
5.2.1. Araştırmadan Çıkarılan Öneriler.....	72
5.2.2. Yapılacak Başka Araştırmalar için Öneriler .....	72
KAYNAKÇA.....	74
EKLER DİZİNİ .....	80
EK 1. ETİK KURUL ONAY BİLDİRİMİ .....	82
EK 2. BİLGİ FORMU .....	83

EK 3. OTİZM SOSYAL BECERİLER PROFİLİ .....	84
(SCOTT BELLINI VE ANDREA HOPF) .....	84
EK 4. GÖNÜLLÜ KATILIM FORMU .....	86
EK 5. VELİ ONAY FORMU .....	87
EK 6. ORJİNALLİK RAPORU .....	88
ÖZGEÇMİŞ. ....	89

## TABLolar DİZİNİ

Tablo 1.1. İki Yüzeyli Tesadüfi Desen için Varyans Bileşenlerinin Kestirilmesi .....	31
Tablo 3.1: Öğrenci Betimsel Bilgileri ve Frekansları .....	47
Tablo 4.1. Elde Edilen Puanların 5 Puanlayıcıya İlişkin Betimsel İstatistikleri .....	52
Tablo 4.2. OSBP Ölçeği ile Yapılan Puanlamalara Ait Cronbach Alfa ( $\alpha$ ) Değerleri	53
Tablo 4.3. 5 Puanlayıcının 15 Maddeye Verdikleri Puanlar Arasındaki Korelasyon Katsayıları .....	54
Tablo 4. 4. b x m x p Desenine Ait G Çalışması Sonucunda Kestirilen Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri .....	56
Tablo 4.5. b x m x p Desenine Ait Karar Çalışması ile Puanlayıcı ve Madde Sayılarının Artırılıp Azaltılmasıyla Yapılan Senaryolara Göre G ve Phi Katsayıları .....	58
Tablo 4.6. (p : b) x m Desenine Ait G çalışması Sonucunda Kestirilen Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri .....	60
Tablo 4.7. (p : b) x m Desenine Ait Karar Çalışması ile Puanlayıcı ve Madde Sayılarının Artırılıp Azaltılmasıyla Yapılan Senaryolara Göre G ve Phi Katsayıları .....	61
Tablo 4.8. b x m x p ve (p : b) x m Desenlerine Ait G çalışması Sonucunda Kestirilen Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri	63
Tablo 4.9. b x m x p ve (p : b) x m Desenlerine Ait Karar Çalışması ile Puanlayıcı ve Madde Sayılarının Artırılıp Azaltılmasıyla Tasarlanan Senaryolardan Elde Edilen G ve Phi Katsayıları .....	65

## ŞEKİLLER DİZİNİ

Şekil 1.1. Genellenebilirlik Kuramının Kökeni ve Kavramsal Çerçevesi (Brennan, 2001) .....	26
Şekil 1.2. İki Yüzeyle Çaprazlanmış b x m x p Deseni İçin Varyans Bileşenleri .....	32
Şekil 1.3. Tesadüfi b x m x p Desende Mutlak Değerlendirmeler İçin Mutlak Hata Kaynağı .....	33
Şekil 1.4. Tesadüfi b x m x p Desende Bağlı Değerlendirmeler İçin Bağlı Hata Kaynağı .....	34

## SİMGELER VE KISALTMALAR DİZİNİ

**OSBP:** Otizm Sosyal Beceriler Profili

**SPSS:** Statistical Package for Social Sciences

**KTK:** Klasik Test Kuramı

**GK:** Genellenebilirlik Kuramı

**G Çalışması:** Genellenebilirlik Çalışması

**K Çalışması:** Karar Çalışması

**G katsayısı:** Genellenebilirlik Katsayısı

**Phi Katsayısı:** Güvenirlilik Katsayısı

**b:** birey

**m:** madde

**p:** puanlayıcı

**b x m:** birey x madde ortak etkisi

**b x p:** birey x puanlayıcı ortak etkisi

**m x p:** madde x puanlayıcı ortak etkisi

**b x m x p:** birey x madde x puanlayıcı artık etkisi

**p : b:** puanlayıcı : birey ana etkisi

**(p : b) x m:** puanlayıcı : birey x madde artık etkisi

## 1. GİRİŞ

Bir bilim dalı olarak ölçme, günlük yaşamda ve bilimde oldukça önemli bir yer tutmaktadır. Günlük hayatımızın her alanında ölçme, ölçme birimi ve ölçme araçları ile karşımıza çıkmaktadır. Bilimde ise ölçme, bilimsel çalışmaların her aşamasında karşımıza çıkmaktadır. Bir bilim dalındaki bilimsel çalışmaların uygulamaya konulması, bu çalışmaların işlevselliğinin belirlenmesi, aynı şekilde bu çalışmaların geliştirilmesi için gerekli olan güvenilir ölçme araç ve yöntemlerin bulunmasında ölçmeden yararlanılmaktadır (Baykul, 2000).

Ölçme, olay veya varlıkların belli bir nicel veya nitel özelliğe sahip olup olmadıklarını, sahiplerse ne ölçüde sahip olduklarını belirleme işlemi olarak ifade edilmektedir (Özçelik, 1991). Weitzenhoffer (1951) ölçmeyi bir gözlemci tarafından fiziksel olaylar üzerinde yapılan bir gözlem olarak tanımlarken, Stevens (1946) ölçmeyi bazı kurallara göre nesne ya da olayları numaralandırmak olarak tanımlamışlardır. Lord ve Novick (1968) ve Torgerson (1958), Stevens'in tanımına ek olarak ölçmenin, nesnenin kendisine değil sahip olduğu bir özelliğe uygulanabileceğini vurgulamışlardır (Crocker ve Algina, 1986). Örneğin ölçme; masanın kendisine değil boy uzunluğuna, havanın kendisine değil sıcaklığına, karpuzun kendisine değil kütlesine veya hacmine uygulanabilmektedir (Turgut ve Baykul, 2014).

Bir kimyager bir karışımın kütlesini ölçerken ya da bir biyolog bir su damlasındaki bakteri sayısını hesaplarken, aslında yapılan işlem inceleme altındaki nesnelerin belirli bazı özelliklerinin ölçülmesidir. Aynı şekilde bir eğitim psikoloğu çocuğun kendisini değil, boyu, kilosu gibi fiziksel özelliklerini veya sosyal olgunluk, sosyal beceri gibi psikolojik özelliklerini ölçer. Fiziksel özelliklerin aksine, psikolojik özellikler doğrudan ölçülemez (Crocker ve Algina, 1986). Psikolojik özelliklerin ölçülmesinde değişkenler, fiziksel büyüklükler değil, zeka, başarı, kaygı, tutum ve beceri gibi niteliklerdir (Baykul, 2000).

### 1.1. Sosyal Beceriler

Combs ve Salaby (1977)'e göre sosyal beceriler, sosyal bir ortamda kabul edilebilir bir şekilde başkalarıyla iletişime geçebilme yeteneğidir. Bu iletişim bazen kişinin kendisine, bazen karşılıklı olarak bazen de sadece başkalarına yarar

sağlayacak şekilde olur. Libert ve Lewinsohn (1979) ise sosyal becerileri, olumlu ve olumsuz pekiştirilen davranışların ortaya çıkmasını sağlayan, cezalandırılan ya da başkaları tarafından sönmesine neden olunan davranışların oluşmasını engelleyen karmaşık beceriler olarak tanımlamışlardır (aktaran Gresham, 1986).

Bu alanda çalışan araştırmacılar video kayıtlarında ya da rol model oyununda sosyal etkileşimleri (göz teması süresi, konuşma sırasında göz teması kurma, yüz ifadelerini vb.) değerlendirir ve sosyal becerilerle ilgili izlenimsel değerlendirmeler yaparlar. Bu sebeple sosyal beceriler doğrudan gözlenemeyeceği için genellikle daha geniş bir kavram olan sosyal davranışların altında değerlendirilir (Gresham, 1986). La Grace (1993) de sosyal beceriyi olumlu sosyal etkileşimlerin başlamasını ve devam etmesini sağlayan olumlu sosyal davranışlar olarak tanımlamıştır.

Sosyal yeterlilik ve sosyal beceri terimleri alanyazında sık sık birbiriyle karıştırılıp aynı anlamda kullanılsa da iki kavram birbirinden farklıdır. Sosyal beceriler, toplumsal yaşamda hayatta kalabilmek ve kaliteli bir yaşam sürebilmek için gerekli olan bilişsel, sözel ve sözel olmayan davranışlardır. Bu beceriler problem çözme kapasitesini, gelen sosyal mesajların doğru yorumlanabilmesini ve mesaj gönderebilme becerilerini kapsamaktadır. Bireyin sosyal becerilerinin çevresindeki insanlar üzerindeki etkileri ise o bireyin sosyal yeterliliğini oluşturmaktadır (Lieberman, Mueser, Wallace, Jacobs, Eckman ve Massel, 1986). Bir anlamda sosyal yeterlilik, sosyal becerileri kapsamaktadır (Merrel ve Cardella, 1999).

Mc Fall (1989)'a göre sosyal beceriler, toplum tarafından yapılması öngörülen görev ve rolleri yerine getirebilmek için birey tarafından sergilenen belli davranışlardır. Öte yandan sosyal yeterlilik ise başkaları tarafından bireyin görev ve rollerini yeterli düzeyde yapıp yapmadığına dair yargılamaya dayanan değerlendirmelerdir. Bu yargılar genellikle çevredeki insanların düşüncelerine ya da belli bir norm grubuyla karşılaştırma yapılmasına dayanır (aktaran Avcıoğlu, 2001).

Otizm Spektrum Bozukluğu (OSB), sosyal etkileşimlerde yüksek seviyede yetersizlikler ve stereotip davranışlar ile karakterize nöro-gelişimsel bir bozukluktur (Özdemir, Diken, İ. H., Diken, Ö. ve Şekercioğlu, 2013). Bu grupta yer alan bireylerde sosyal etkileşimde zayıflık (sosyal ve duygusal karşılıklılık eksikliği vb.),

iletişim sınırlılıkları (konuşma becerilerindeki yetersizlikler vb.), stereotipik davranışlar (kendini tekrar eden motor hareketler, rutinlere olan bağlılık vb.) görülmektedir (Heward, 2012).

Otizimli bireylerin iletişim becerilerindeki sınırlılıkları, yaşına göre yeterli dil becerilerini gösterememek, konuşurken alışılmadık sesler çıkarmak, konuşma başlatma, sürdürme ve sonlandırmada başarısız olmak, konuşma konularında seçici olmak (sadece favori konularında konuşmak) şeklinde sıralanabilmektedir. Bunlara ek olarak sadece kendisine anlam ifade eden kelimeleri tekrar etme, didaktik konuşma özelliklerine sahip olma, sosyal oyunlara karşı ilgisiz olma, sembolik ve dramatik oyun oynamada yetersiz olma da iletişim becerilerindeki yetersizlikler arasında yer almaktadır. Sosyal etkileşim için gerekli sözel olmayan davranışlarda hiç göz teması kuramamak, birden bire gözlerini karşısındakinin gözlerine dikmek ve kaçırmak, konuşurken jest ve mimik kuramamak gibi yetersizlikler gösterebilmektedirler. Ayrıca başkalarına yaklaşımda sıra dışı özellikler gösterme (nerede duracağını bilememek), , arkadaşlık kurmada sıkıntı yaşama, grup etkinliklerinde zorlanma, yalnızlığı yeğleme, başkalarının dikkatini çekme çabası göstermeme, sözel övgüler karşısında tepkisiz kalma, başkalarına karşı ilgisiz olma, başkalarının duyguları anlamada yetersiz olma konularında da sınırlılıkları bulunmaktadır. Bunlara ek olarak sıra dışı beden hareketleri sergilemek (kendi eksenini etrafında dönme vb.), sıra dışı el hareketleri sergilemek (ellerini sürekli çırpma vb.), eline aldığı nesnelere duyuşsal özellikleriyle fazla ilgilenmek (koklamak, sürekli evirip çevirmek vb.), hareket eden nesnelere aşırı ilgi göstermek gibi stereotipik davranışlar sergilemektedirler (Kırcaali-İftar, 2012).

İletişim ve sosyal etkileşim alanındaki sınırlılıklar çocuklar için çeşitli olumsuz sonuçlar doğurmaktadır. Örneğin, bir çocuktan okula başladığı zaman, akademik ve sosyal görevlerini yerine getirecek belli becerileri (bazı basit istekleri yerine getirme, problem çözme becerileri, grup çalışmalarına katılma vb. ) göstermesi beklenmektedir (Lane, Givner ve Pierson, 2004). Bu becerileri yerine getiremeyen çocuğun, öğretmen ve akranları tarafından kabulü zorlaşmakta, arkadaşları tarafından istenmemesine neden olmaktadır. Yetersizlikleri olan pek çok çocuk, arkadaşlarıyla iletişim başlatma ve sürdürmede sıkıntı çekmekte, iletişimi yorumlamada hatalar yapmakta, grup çalışmalarına katılamamaktadır (Avcioğlu, 2011). Bu sebeple sosyal becerilerde sınırlılıkları olan çocukların akranlarına



oranla kendilerini daha çok yalnız hissetmektedirler (Bauminger, Schulman ve Agam, 2003). Ayrıca, sosyal yeterlilik ile akademik beceriler arasında oldukça yüksek bir ilişki bulunmaktadır. Sosyal beceri yetersizlikleri nedeniyle sınıf ortamından dışlanmış çocuklar, akademik başarı göstermekte sıkıntı çekmekte ve geri planda kalmaktadırlar (Bursuch ve Asher, 1986). Yüksek işlevli otizmli çocukların özellikle ergenlik dönemine girdikten sonra akranlarıyla sosyal etkileşime girme istekleri artmaktadır. Ancak sosyal beceri yetersizlikleri arkadaşlık kurmalarına engel olmaktadır. Ortama uyum sağlamanın önemini fark ettikleri an, kendilerini dışlanmış ve izole edilmiş hissetmektedirler. Bu da özgüven eksiklikleri, depresyon, kaygı düzeylerinde yükseklik vb. gibi sonuçlar doğurmaktadır (Tse, Strulovitch, Tagalakis ve Fombonne, 2007). Yeterli sosyal beceri seviyesine sahip olmayan çocuklar, yetişkinlik dönemlerinde de toplumda ve iş hayatında sıkıntı yaşamaktadırlar (Hall ve Schlesinger, 1997).

Pek çok sosyal beceri sistematik olmayan bir şekilde arkadaş grupları ve aile içerisinde öğrenilmektedir. Otizmli çocuklar iletişim kurmada sıkıntı yaşamaları sebebiyle aile ve akranlarından kopuk bir yaşam sürmektedirler. Bu sebeple sosyal becerileri doğal ortamda gözlemleyerek arkadaş ve ailelerinden öğrenememektedirler. Bu becerilerin öğrenilebilmesi için çocuğun özelliklerine göre yapılandırılmış sosyal beceri öğretim programlarına ihtiyaç duyulmaktadır (Hall ve Schlesinger, 1997).

Bireylerin günlük yaşam gerekliliklerini yani kişisel ve işlevsel yeterliliklerini yerine getirebilmeleri için üç temel öge gerekmektedir. Bunlar bilgi edinimi, ilgili becerilerin gelişimi, bilgi ve becerilerin farklı durum ve ortamlara uygulanmasıdır (Pollaway, Serna, Patton ve Bailey, 2014). Parsons (2006)'a göre sosyal beceri uygulamalarının çocuğun tüm hayatına genellenmesi oldukça önemlidir. Genelleştirme problemini azaltmanın bir yolu, çocuğun akran grubuyla birlikte yer alabileceği sosyal beceriler öğretim programının hazırlanmasıdır (Choi ve Heckenlaible, 1998).

Sosyal beceri öğretim programlarının etkinliğinin artırılması için çocukların gereksinimlerinin ve yetersizliklerinin farklı değerlendirme yöntemleri kullanılarak belirlenmesi gerekmektedir (Sucuoğlu ve Özokçu, 2005).

## 1.2. Sosyal Becerilerin Değerlendirilmesi

Sosyal beceri öğretim programlarının temelini, çocukların sosyal becerilerinin değerlendirilmesi oluşturmaktadır. Değerlendirme yapılırken çocuğun özelliklerine ve değerlendirmenin amacına göre uygun yöntemin seçilmesi oldukça önemli bir rol oynamaktadır. Sosyal becerilerin ölçülmesinde 5 temel yöntem kullanılmaktadır. Bunlar: görüşme, gözlem, öz-değerlendirme, sosyometri ve dereceleme ölçekleridir. Bütün bu yöntemlerin kendilerine has avantaj ve dezavantajları bulunmaktadır. Her yöntem kendi amacına hizmet etmek noktasında geçerli olmasına rağmen, sosyal becerileri değerlendirme de hepsi aynı oranda kullanışlı değildir (Merrel, 2001).

### 1.2.1. Görüşme

Görüşme, çocuğun sıkça vakit geçirdiği anne/baba, bakıcılar ve öğretmenler gibi çocuk için önemli olan kişilerle niteliksel bilgi toplamak amacıyla yapılmaktadır. Bu görüşmeler sonucunda çocuğun sosyal becerilerdeki eksiklikleri ve gereksinimleri belirlenip, ulaşılmak istenen hedef beceriler belirlenmektedir. Eliot ve Gresham (1987)'e göre görüşmeler aşağıdaki durumlar açısından avantaj sağlamaktadırlar;

- a) Sosyal becerilerin gözlenebilir terimlerle tanımlanması
- b) Ölçülmek istenen becerilerin oluşmasını sağlayan öncesindeki ve sonrasındaki durumların belirlenmesi
- c) Hedeflenen becerilerin ölçülmesi için gerekli gözlem ortamının hazırlanması

Görüşmeler, sosyal becerilerin değerlendirilmesinde sıklıkla kullanılan yöntemlerden biridir (Crozier ve Tincani, 2005). Ancak, elde edilen bilgilerin nesnel değil de öznel bilgi olması açısından bu yöntem tek başına çok da güvenilir değildir. Bu sebeple, görüşmeler sonucunda elde edilen bilgilerin, diğer yöntemlerle elde edilen bilgilerle birlikte değerlendirilmesi daha yararlıdır.

Görüşmeler, açık uçlu soruların sorulduğu serbest görüşme ve kapalı uçlu soruların sorulduğu yapılandırılmış görüşme olmak üzere ikiye ayrılmaktadır. Detaylı bilgi edinmek açısından serbest görüşmeler daha avantajlı olurken, konunun dağılmaması ve istenilen bilgilere doğrudan ulaşılabilmesi açısından yapılandırılmış görüşmeler daha avantajlı olmaktadır. Bu sebeple görüşmenin amacı ve görüşülen kişilerin tutumuna göre sadece serbest, sadece

yapılandırılmış ya da yarı yapılandırılmış bir görüşme uygulanabilmektedir (Avcıoğlu, 2011).

### **1.2.2. Öz-değerlendirme**

Adından da anlaşılacağı gibi öz değerlendirme, çocuğun kendi kendini değerlendirdiği bir yöntemdir. Çocuktan kendi sosyal becerilerini değerlendirmesi beklenmektedir. Öz değerlendirme yapılırken cümle tamamlama, resim çizme gibi çeşitli araçlar kullanılmaktadır. Toplanan bilgiler daha sonra norm grubu olan diğer akranlarından elde edilen bilgilerle karşılaştırılmaktadır (Demir, 2012).

Sosyal becerilerin değerlendirilmesinde diğer yöntemler kadar sık kullanılsa da çocukların sosyal becerileri hakkında ek bilgi toplanması açısından faydalı bir tekniktir (Eliot ve Gresham, 1987). Öz-değerlendirme yönteminin en büyük dezavantajı ise öznel bilgi elde edilmesidir. Çocukların kendilerini değerlendirirken objektif olamamaları ya da kendi sosyal beceri yetersizliklerinin farkında olmamaları öz değerlendirmenin geçerlik ve güvenilirliğini düşürmektedir. Bu sebeple, sosyal beceri yetersizliklerinin belirlenmesinde bu yöntem tek başına kullanılamamaktadır (Merrell, 2001)

### **1.2.3. Sosyometri**

Sosyometri, özellikle sosyal beceri yetersizliği olan çocukların akranlarıyla olan ilişkilerinin belirlenip, ona göre sosyal beceri öğretimi programlarının hazırlanmasında etkili bir yöntemdir. (Asher, Singleton, Tinsley ve Hynel, 1979). Bu yöntem ile bir grup içerisindeki çocukların hangilerinin akranları tarafından kabul edildiği ve hangilerinin dışlandığı belirlenebilmektedir. Örneğin, en çok kiminle oynamayı sevdikleri, kimin yanına oturmak istedikleri, en iyi anlaştıkları ve ya anlaşamadıkları arkadaşlarının kim olduğuna dair sorular sorularak sınıf içindeki akran ilişkileri değerlendirilebilmektedir. Sosyometrik ölçme tekniklerinin akran atamaları, akran derecelmeleri ve akran değerlendirmeleri olmak üzere üç tipi bulunmaktadır (Eliot ve Gresham, 1987). Merrell (2001)'e göre en büyük avantajı sosyal beceri yetersizliği olan çocukların akranları tarafından dışlanıp dışlanmadığının belirlenmesini sağlayan sosyometri yöntemi ile bu çocukların hangi becerilerde güçlük çektiği belirlenmemektedir. Bu da bu yöntemin en büyük sınırlılığıdır.

#### 1.2.4. Gözlem

Eliot ve Gresham (1987)'e göre gözlem, sosyal becerilerin doğal ortamda ve gerçekleştiği sırada gözlemlenip kaydedilmesi demektir. Görüşme yönteminde olduğu gibi amaç çocuğun sergilediği sosyal becerilerle ilgili duygu, düşünce ve görüşlerin alınması değil, çocuğun davranışlarının incelenmesidir (Avcioğlu, 2011). Bu süreçte çocukların akranlarıyla ve yetişkinlerle olan iletişimleri, bu sırada sergiledikleri davranışları uygun bir teknik ile gözlemlenmektedir. Tanımı gereği doğal ortamda gözlem 3 ana bileşenden oluşmaktadır (Jones, Reid ve Patterson, 1979).

- 1) Doğal ortamda davranışın gerçekleştiği sırada gözleme ve kayıt altına alınması
- 2) Objektif ve bu konuda uzman gözlemciler tarafından gerçekleştirilmesi
- 3) Tüm gözlemcilerin hem fikir olacağı şekilde davranışın tanımlanmasıdır.

Gözlem sadece sosyal becerilerdeki yetersizlikleri belirlemek amacıyla kullanılmaz, aynı zamanda bu becerilerin takibi ve sosyal beceri öğretimi programını geliştirme amacıyla da kullanılmaktadır. Bu sebeple sık sık kullanılan bir yöntemdir (Aldred, Green ve Adams, 2004; Lane, Givner ve Person, 2004). Doğal ortamda gözlemin yanı sıra yapılandırılmış ortamlarda da gözlem yapılabilmektedir. Yapılandırılmış gözlemlerde, doğal ortam koşulları bir laboratuvar ya da klinikte sağlanır. Ancak Eliot ve Gresham (1987)'e göre sosyal becerilerinin değerlendirilmesinde en geçerli yöntem çocukların davranışlarının okul, sınıf, oyun alanları gibi doğal ortamlarında gözlemlemektir.

Doğal ortamda sosyal becerilerin değerlendirilmesi için gözlem yapılırken izlenilmesi gereken aşamalar vardır. İlk olarak değerlendirilmek istenen beceri belirlenir ve bu beceri gözlenebilir ve ölçülebilir şekilde tanımlanır. İkinci basamakta Merrell (2001)'e göre gözlem yapılırken, gözlemcinin hangi kayıt tekniğini kullanacağına karar vermesi gerekmektedir. Değerlendirme uzmanları genellikle olay kaydı, aralık kaydı, aralık-zaman örnekleme, süre kaydı ve bekleme kaydı olmak üzere 5 farklı kayıt tekniğini kullanmaktadırlar.

- 1) Olay kaydı: Gözlem süresince belli bir davranışın kaç defa oluştuğunu belirtmektedir.

- 2) Aralık kaydı: Gözlem süresi belli aralıklara bölünüp, belli bir davranışın bu aralıklarda kaç defa olduğu gözlemlenmektedir. Gözlemin amacına ve becerinin özelliğine göre tüm aralık kaydı ve yarı-kısmi aralık kaydı olarak 2 çeşit aralık kaydı bulunmaktadır.
- 3) Aralık zaman örnekleme: Aralık zaman kaydına benzemektedir. Ancak aralık zaman örneklemeinde davranış, belirlenmiş zaman aralığından hemen sonra kaydedilmektedir.
- 4) Süre kaydı: Belli bir davranışın ne kadar sürdüğü kaydedilmektedir.
- 5) Bekleme kaydı: Bir davranışın bitmesinden sonra ikinci davranışın başlamasına kadar geçen süre belirlenmektedir.

Sosyal becerilerin doğal ortam gözleminde yukarıda bahsedilen 5 kayıt türü kullanılmasına rağmen, araştırmacılar tarafından olay kaydı ve aralık kaydı daha çok tercih edilmektedir. Kayıt türüne karar verildikten sonra gözlemin ne zaman ve nerede yapılacağı belirlenmesi gerekmektedir. Gözlem yapıp veriler elde edildikten sonra, bu verilerin çocuğun akranlarından elde edilen verilerle karşılaştırılıp yorumlanması bu sürecin son basamağını oluşturmaktadır (Lane, Menzies, Barton, Doukas ve Munton, 2005)

Sosyal becerilerin değerlendirilmesinde gözlem tekniğinin kullanılmasının olumlu yanları bulunmaktadır. Çocuğun doğal ortamda gözlenmesi ile diğer değerlendirme yöntemlerinden elde edilemeyecek esas bilgiler elde edilmektedir. Görüşme ve öz-değerlendirme yöntemlerinden elde edilen duygu, düşünce ve görüşlerin, gözlem yaparak elde edilen bilgilerle değerlendirilmesi daha geçerli ve güvenilir sonuçlar sağlamaktadır. Fazlaca araç ve gereç kullanımı gerektirmediği için ekonomik bir tekniktir (Avcıoğlu, 2011).

Avantajlarının yanı sıra her teknikte olduğu gibi gözlem tekniğinin de bazı sınırlılıkları bulunmaktadır. Bir gözlemin güvenilirliği gözlemcinin ön yargılarına, gözlem yeteneğine, sosyal becerileri anlama ve algılama gücüne, olaylara bakış açısına vb. bağlıdır. Bu sebeple gözlem yapılmadan önce gözlemcilerin iyi bir eğitimden geçmesi gerekmektedir. Bu eğitimler zaman alıcı ve maliyetlidir. Ayrıca, davranışlar hakkında yeterli bilgi alabilmek için gözlemlerin birden fazla ortamda ve zamanda yapılması gerektiğinden zaman alıcı ve yorucu olabilmektedir (Merrel, 2001; Demir, 2012).

### 1.2.5. Dereceleme Ölçekleri

Çocuğun kendisi ya da çocuğa yakın olan öğretmen, ebeveyn veya akranları tarafından doldurulan, çocuğun sergilediği her bir davranışı ne kadar sergilediği ya da her bir tanıma ne kadar uyduğunu belirlemek amacıyla pek çok davranışı ya da tasvire dayalı maddeleri içeren ölçeklere dereceleme ölçeği denir. Dereceleme ölçekleri çocukların davranışlarının, cevaplarının ve performanslarının belirli bir ölçüte göre puanlanmasında kullanılan bir klavuzdur (Turgut ve Baykul, 2014).

Öğrencilerin sosyal becerilerinin detaylı olarak değerlendirilmesini sağlayan dereceleme sistemi ile aynı zamanda bu becerilerin öğrenimi ve kullanımını etkileyen problem davranışlar da ölçülmektedir (Sucuoğlu ve Özokçu, 2005). Öğretmenler tarafından yapılan dereceleme yöntemi, hangi sosyal becerilerde yetersizliklerin olduğunun belirlenmesi ve bu becerilere sosyal beceri öğretimi programında ağırlık verilmesi açısından önemlidir (Eliot ve Gresham, 1987).

Dereceleme ölçeklerinin diğer tekniklere göre birçok avantajı vardır. Bunlar;

- 1) Doğal ortamda gözlem tekniğine göre çok zaman almaması ve gözlemci eğitiminin süresinin çok uzun olmaması
- 2) Doğal ortamda gözlemlenemeyecek kısa süreli ve çok sık ortaya çıkmayan davranışlar hakkında bilgi sağlaması
- 3) Görüşme ve öz değerlendirmelere oranla daha güvenilir bilgiler vermesi
- 4) Kendi davranışıyla bilgi veremeyecek küçük çocuklarda ya da kendini ifade edemeyen özel gereksinimle çocuklarda uygulanabilmesi
- 5) Çocuğun doğal ortamda gözlemlenmesi sonucu doldurulması
- 6) Değerlendirmenin çocuğu iyi tanıyan aile, öğretmen ya da akranları tarafından doldurulmasıdır.

Yukarıda belirtilen avantajları sayesinde dereceleme ölçekleri kısa sürede, uygun maliyetle güvenilir sonuçlar vermektedir. Bu sebeple dereceleme ölçekleri çocuğun davranışlarının değerlendirilmesinde en çok kullanılan yöntemdir (Merrell, 2001).

Dereceleme ölçekleriyle değerlendirmeye, geçerlik ve güvenilirliğini tehlikeye sokan bazı hatalar karışmaktadır (Turgut ve Baykul, 2014). Bunlar:

- 1) Dikkatsizlik: Dereceleme ölçeğindeki ifadelerin tam açık olmamasından ya da davranışın tam olarak algılanamamasından kaynaklanabilir.
- 2) Kişisel yanlılık: Ölçeği dolduran kişinin çocuğa karşı olan subjektif eğilimlerinden kaynaklanır. Özellikle ebeveynler çocuklarını değerlendirirken gerçekleri farkında olmadan çarpıtabilirler.
- 3) Cömertlik/Eli sıkılık: Değerlendirmeyi yapan kişinin çocuğa karşı fazla cömert ya da katı olmasından kaynaklanır. Bazen ölçeği dolduran kişiler olumsuz değerlendirme yapmamak için hep yüksek puanlar vermeye eğilimlidirler.
- 4) Hareleme etkisi: İlk boyutta iyi değerlendirme yaptıktan sonra hep iyiye doğru ya da düşük değerlendirme yaptıktan sonra hep düşük puan verme eğiliminden kaynaklanır.
- 5) Merkeze kayma etkisi: Değerlendirmeyi yapan kişinin uçlarda yer alan “her zaman” , “hiçbir zaman” gibi dereceleri kullanmaktansa ortalarda kalan dereceleri kullanmaya eğiliminden kaynaklanır.
- 6) Genelleme hatası (halo effect): Çocuğun dereceleme ölçeğiyle değerlendirilen davranışına göre değil de, iyi ya da kötü özelliklerine yani farklı bir niteliğe göre genelleme yapılarak değerlendirilmesidir (Yıldırım, 2014).
- 7) Gözlem yetersizliği: Davranışın dikkatlice ya da yeterince gözlemlenememesinden dolayı gelişigüzel olarak ölçeğin işaretlenmesinden kaynaklanır.

Gözlem sonuçlarını sayısal verilere çevirmeye yarayan dereceleme ölçeklerinde izlenen süreç “sıralama”dır. Dereceleme ölçekleri kullanılarak yapılan değerlendirme, ölçeği dolduran kişinin ölçekte bulunan her bir maddeye ilişkin cevabını, birbirini mantıksal bir sırayla izleyen cevap seçenekleri (ölçek noktaları) üzerinde kendisine en yakın gelen seçeneği işaretlemesi ile gerçekleşir (Büyüköztürk, 2013). Dereceleme ölçeklerinin sayısal derecelendirme ölçeği, grafiksel derecelendirme ölçeği, betimsel derecelendirme ölçeği ve karşılaştırmalı derecelendirme ölçeği olmak üzere 4 çeşidi bulunmaktadır (Avcıoğlu, 2011).

### 1.2.5.1. Sayısal Derecelendirme Ölçeği

Gözlenen çocuğun derecelendirilmek istenen beceriye sahip olup olmadığını, sahipse ne derece de sahip olduğunu belirlemek amacıyla kullanılan ölçeklerdir. Ölçek noktaları 5-11 arasında olabilir. Ölçek noktalarının her biri kısa bir ifade ile tanımlanır. Örneğin 1 rakamı, “hiçbir zaman” kelimesinin yerini tutmakta, 2 rakamı “bazen”, 3 rakamı “sık sık”, 4 rakamı “her zaman” ifadelerinin yerini tutar.

Örnek: Gözlemlerinizi doğrultusunda uygun olduğunu düşündüğünüz dereceye karar veriniz ve yuvarlak içine alınız.

<b>Sosyal Beceriler</b>	<b>Hiçbir zaman</b>	<b>Bazen</b>	<b>Sık sık</b>	<b>Her zaman</b>
1. Oyun ve etkinlikler esnasında sıra alır.	1	2	3	4
2. Bir kişi hakkında sorular sorar.	1	2	3	4
3. Konular hakkında sorular sorar.	1	2	3	4

### 1.2.5.2. Grafikselle Derecelendirme Ölçeği

Ölçülecek nitelik ya da davranışın, bir doğru üzerinde dereceli olarak gösterilmesiyle oluşturulmuş ölçeklere grafikselle derecelendirme ölçekleri denir. Bu doğru çizgi yatay ise yatay grafikselle derecelendirme ölçekleri, çizgi dikey ise dikey grafikselle derecelendirme ölçeği olarak adlandırılır.

#### 1.2.5.2.1. Yatay Grafikselle Derecelendirme Ölçeği

Bu ölçekte ölçülecek nitelik için yatay bir doğru dan yararlanılır. Yatay doğru uzunlukları standart olan birimlere ayrılır. Doğrudaki uç birimlerin üzerine niteliğin derecesini gösteren sayı, sözcük veya ifadeler konulur. Bu birimlerin görevi ölçeği dolduran kişiye süresiz kategoriler sağlamak değil, onun cevabını yerleştirmesine rehberlik etmektir. Doğru kullanım için uçlardaki ifadelerin aşırı olmamasına dikkat edilmelidir (Balci, 2001). Derecelendirmeyi yapan kişi doğru üzerindeki rakamları daire yuvarlak içine alır ya da o birime bir işaret koyar.



Örnek:

1. Kişisel temizliği

Çok iyi				İyi değil
5	4	3	2	1

2. Birebir akran iletişimleri

Çok iyi				İyi değil
5	4	3	2	1

**1.2.5.2.2. Dikey Grafiksel Derecelendirme Ölçeği**

Bu ölçekte gruptaki herkes tek bir nitelik yönünden grup üyeleriyle karşılaştırılarak değerlendirilir.

Örnek: Kişinin arkadaşlık ilişkileri yönünden durumunu uygun yüzdeliğin karşısına ve adının bulunduğu dikey aralık içine (X) işareti koyarak belirtiniz.

%	Değerlendirilecek Kişiler					%
	Cahit	Cemile	Zeynep	Kadir	Kübra	
100						100
90			X		X	90
80				X		80
70		X				70
60						60
50	X					50
40						40
30						30
20						20
10						10
0						0

Örnekte Zeynep'in arkadaşlık ilişkileri %90, Cemile'ninki ise %70 olarak belirtilmiştir.

### 1.2.5.3. Betimsel Derecelendirme Ölçeği

Derecelendirilmesi istenen niteliğin betimsel ifadeler kullanılarak derecelendirilmesiyle oluşan ölçeklerdir. Niteliğin farklı dereceleri için farklı tanımlamalar yapılır.

Örnek: Bireyin akranlarıyla etkinliklere katılma durumuna uygun düşen ifadenin önündeki parantezin içine (X) işareti koyunuz.

( ) Akranlarıyla etkinliklere katılmayı çok sever

( ) Teklif edildiği zaman akranlarıyla etkinliklere katılır

( ) Akranlarıyla etkinliklere katılma ya da yalnız takılma konusunda bir tercihi yoktur.

( ) İlişki kurmada güçlük çeker, yalnız takılmayı daha çok tercih eder

( ) Akranlarıyla etkinliklere asla katılmaz

Betimsel derecelendirme ölçekleri sadece betimsel olabileceği gibi betimsel grafiksel derecelendirme ölçeği şeklinde de olabilir. Betimsel grafiksel derecelendirme ölçeği yatay grafiksel derecelendirme ölçeğine benzer. Tek farkı her bir derece hem sayı ile hem de betimsel ifade ile belirtilir.

Örnek: Grup içerisinde sorumluluk alma durumuna uygun olduğunu düşündüğünüz dereceye karar veriniz ve (x) işareti koyunuz.

	Diğer grup üyelerine göre daha az alır	Gruptaki üyeler kadar alır	Diğer grup üyelerine göre daha fazla alır	En çok sorumluk alan bireylerdendir
Hiç almaz				
1	2	3	4	5

#### 1.2.5.1. Karşılaştırmalı Derecelendirme Ölçeği

Derecelendirilmesi istenen bir özellik, normal olarak diğer kişilerde olması beklenen derece ya da seviye ile karşılaştırılıp, bu ölçek üzerinde değerlendirilir. Örneğin, liderlik özelliğine sahip en üst ve en alt seviyelerdekideki öğrencilerin belirlenmesi için, bu özelliğe en az sahip, orta seviyede sahip ve en yüksek seviyede sahip öğrenciler belirlenir. Daha sonra herhangi bir öğrenci liderlik özelliği bakımından bu üç öğrenciyle kıyaslanarak “eşit, daha iyi, daha zayıf” gibi yargılara varılır (Balcı, 2001).

#### 1.3. Problem Durumu

Doğası gereği ölçme, belli bir özellik ya da davranış hakkında bilgi edinilmek istendiğinde, ölçme sonuçları içerisinde çeşitli sebeplerden kaynaklanan hata

barındırmaktadır. Ölçmeye bu açıdan bakıldığında; hata, yanlışlık kavramı olarak düşünülmemelidir. Buradaki hata kavramı, sürecin doğal bir parçası olup, bir şekilde süreçte yer almaktadır (Brennan, 2001).

Güvenirlikle ilgili alanyazın incelenip, güvenilirliği etkileyen hataların nereden kaynaklandığı göz önünde bulundurulduğunda, sistematik bir sınıflamaya ulaşılabilir. Bunlar ölçmenin yapıldığı kişiler, ölçmeyi yapan kişi ya da kişiler, ölçmenin yapıldığı koşullar ve ölçme aracıdır (Crocker ve Algina 1986; Baykul 2000; Lord ve Novick 1968).

Puanlayıcıların dâhil olduğu ölçme sürecinde, ölçme aracı ve ölçütün niteliği dışında puanlayıcılar da önemli bir hata kaynağı olmaktadır. Özellikle bireyin kendisi tarafından değil de gözlemlenmesi sonucu başkaları tarafından (aile, öğretmen vb.) değerlendirildiği dereceleme ölçekleri ile karara varma süreci, çoğunlukla puanlayıcının yargısı ile gerçekleşmektedir. Özel eğitim alanında yapılan çalışmalarda çoğunlukla otizmlili bireylerin kendilerini değerlendirecek yeterlilikte olmamaları sebebiyle değerlendirmeler aile, öğretmen gibi bireye yakın kişiler tarafından yapılmaktadır. Bu sebeple, dereceleme ölçekleri ile değerlendirme yapılırken puanlayıcılar arası güvenilirliğin test edilmesi oldukça önem kazanmaktadır.

Puanlayıcılar arası güvenilirliğin belirlenmesinde kullanılan çeşitli kuram ve uygulamaları bulunmaktadır. Ancak bu kuramların gerçek hayatta doğruluklarının ve işlevselliklerinin birbiriyle karşılaştırılıp, test edilmesi gerekmektedir. Doğan (2002) tarafından, bu kuramların artı ve eksilerinin, birbirleriyle benzerlik ve farklılıklarının belirlenmesinin bilimsel çalışmalar için oldukça önemli olduğu vurgulanmaktadır.

Araştırmada klasik test kuramı ve genellenebilirlik kuramının aynı veriler üzerinde uygulanması ve elde edilen parametreler doğrultusunda her iki kuramında avantaj ve dezavantajlarının belirlenmesi, hangi kuramın daha kullanışlı olduğu, hangi kuramdan daha çok bilgi sağlandığının incelenmesi amacıyla belirtilen problem cümlesi ve alt problemlere yanıt aranmaktadır.

#### **1.4. Araştırmanın Amacı Ve Önemi**

Bu araştırmanın amacı, Otizm Sosyal Beceri Profili ile otizmlili çocuk ve gençlerin birden fazla puanlayıcı tarafından puanlanması sonucu elde edilen sonuçlarda

puanlayıcılar arası güvenilirliğin klasik test kuramına ve genellenebilirlik kuramına göre sonuçlarının karşılaştırılmasıdır.

Araştırmada, otizmlili bireylerin puanlayıcılar tarafından puanlanmasıyla elde edilen veriler klasik test kuramına göre incelenip, puanlayıcılar arası güvenilirlik belirlenmiştir. Daha sonra genellenebilirlik kuramı ile puanlayıcıların birlikte ve dönüşümlü puanlanmasıyla oluşturulan desenler incelenmiştir. Genellenebilirlik kuramının kullanımını ve farklı desenlerin kullanılabilirliğini örnekleme amacı da taşımaktadır. Genellenebilirlik kuramındaki farklı desenler için karar çalışmalarında puanlayıcı ve madde sayılarının artırılıp azaltılmasıyla en güvenilir sonuçlar elde etmek için uygun durumlar belirlenmiştir. İki kuramdan elde edilen sonuçlar karşılaştırılarak ölçme durumuna en uygun kuramın belirlenmesi amaçlanmaktadır. Ayrıca bu kuramlardan hangisinin daha çok bilgi sağladığı, kuramların üstünlük ve sınırlılıklarının belirlenmesi ve uygulayıcıların pratikte tercih edebilecekleri kuramın belirlenmesi açısından araştırmanın önemli olduğu düşünülmektedir.

Daha önceden de belirtildiği üzere sosyal beceri yetersizlikleri otizmlili bireylerin sosyal ve eğitim hayatlarında iletişim kurmalarında sıkıntı yaşamalarına sebep olmaktadır. Bu sıkıntıları önlemek adına sosyal beceri öğretim programları geliştirilmiştir. Ancak bu programların uygulanabilmesi için ilk adımda bireylerin sosyal beceri yetersizliklerinin belirlenmesi gerekmektedir. Daha sonrasında ise programların işlevselliklerinin takibi, eksik yanlarının belirlenmesi için ölçme ve değerlendirme çalışmalarının yapılması önemli bulunmaktadır. Sosyal beceri yetersizliklerinin belirlenmesi ve programların takiplerinin yapılması genellikle bireye yakın aile fertlerinin ya da öğretmenlerin bireyi gözlemleyip değerlendirmesi ile gerçekleşmektedir. Ancak puanlama yapan kişilerin düşünceleri, tutumları ve yargıları bu sonuçların güvenilirliğini düşürmektedir. Puanlayıcıların ölçme sürecine katılımıyla, ölçme uygulamalarında puanlayıcılar arası uyum ve güvenilirliğin araştırılması giderek artan eğilimdedir (Atılğan, 2004). Türkiye’de alanyazın incelendiğinde özel eğitim alanında puanlayıcılar arası güvenilirliğe dair bir çalışmaya rastlanmamıştır.

Bu araştırma ile hem kuramlardan elde edilen bilgilerin karşılaştırılmasına, hem gelecek araştırmalara hem de özel eğitim alanında ölçme ve değerlendirme sürecine yönelik bir katkı sağlanabileceği düşünülmektedir.

## 1.5. Problem Cümlesi

Otizm Sosyal Beceriler Profili ölçeğinin sosyal karşılıklılık alt boyutunun birden fazla puanlayıcı tarafından puanlanması sonucu elde edilen puanlayıcılar arası güvenilirliğin, klasik test kuramı ve genellenebilirlik kuramındaki farklı desen çalışmaları sonuçları nasıldır?

### 1.5.1. Alt Problemler

1) Klasik test kuramına göre;

1.1. Otizm Sosyal Beceriler Profiline sosyal karşılıklılık alt boyutundan elde edilen puanların 5 puanlayıcıya ilişkin iç tutarlılık düzeyi nedir?

1.2. Otizm Sosyal Beceriler Profiline sosyal karşılıklılık alt boyutuna ilişkin değerlendirmede 5 farklı puanlayıcı arasındaki tutarlılık derecesi nedir?

2) Genellenebilirlik Kuramına göre; birey (b), madde (m) ve puanlayıcı (p) değişkenlerinin çapraz tasarlandığı  $b \times m \times p$  deseninin sonuçları nasıldır?

2.1.  $b \times m \times p$  deseninin G çalışması sonucunda kestirilen varyans bileşenleri ve toplam varyansı açıklama yüzdeleri nasıldır?

2.2.  $b \times m \times p$  deseninde puanlayıcı ve madde sayılarının artırılıp azaltılmasıyla yapılan karar çalışması sonuçları nasıldır?

3) Genellenebilirlik Kuramına göre; birey (b) ve puanlayıcı (p) değişkenlerinin yuvalanmış, madde (m) değişkeninin ise çapraz tasarlandığı  $(p : b) \times m$  deseninin sonuçları nasıldır?

3.1.  $(p : b) \times m$  deseninin G çalışması sonucunda kestirilen varyans bileşenleri ve toplam varyansı açıklama yüzdeleri nasıldır?

3.2.  $(p : b) \times m$  deseninde puanlayıcı ve madde sayılarının artırılıp azaltılmasıyla yapılan karar çalışması sonuçları nasıldır?

4)  $b \times m \times p$  ve  $(p : b) \times m$  desenlerinden elde edilen G çalışması sonuçlarının karşılaştırılması nasıldır?

5)  $b \times m \times p$  ve  $(p : b) \times m$  desenlerinde puanlayıcı ve madde sayılarının artırılıp azaltılmasıyla yapılan karar çalışmaları sonuçlarının karşılaştırılması nasıldır?

## 1.6 Sayılılar

1) Farklı puanlayıcıların yaptıkları puanlamalar birbirinden bağımsızdır.

## 1.7. Sınırlılıklar

1) Araştırma, 5 gönüllü puanlayıcı ile sınırlıdır.

2) Araştırma, Otizm Sosyal Beceriler Profiline sosyal karşılıklılık alt boyutunda elde edilen verilerle sınırlıdır.

3) Araştırma, 6-17 yaş arası 50 otizmlilik çocuk ve genç ile sınırlıdır.

## 1.8. Araştırmanın Kuramsal Temeli

Bu çalışmada puanlayıcılar arası güvenilirlik, ölçmenin temel kuramlarından olan Klasik Test Kuramı ve Genellenebilirlik kuramı ile çalışılmaktadır.

### 1.8.1. Klasik Test Kuramı

Klasik test kuramının temeli 1907-1913 yılları arasında Spearman Brown tarafından atılmıştır. Daha sonra bu kuram, özellikle Guilford(1936), Gulliksen (1950), Magnusson (1967), Lord ve Novick (1968) tarafından ele alınmış ve yeniden ifade edilmiştir (Crocker & Algina, 1986).

Baykul (2000)'e göre güvenilirlik, ölçme sonuçlarının içinde barındırdığı hata puanlarının azlığı şeklinde tanımlanmaktadır. Günümüze kadar psikoloji ve eğitimde ölçme sonuçlarının güvenilirliği üzerine birçok tanım yapılmıştır. Bu tanımlar doğrultusunda güvenilirlik hesaplamalarında klasik test kuramı adres olarak gösterilmiştir. Klasik test kuramına göre, Gözlenen puan (X); gerçek puan (T) ve rastgele hatadan (E) oluşmaktadır. Buna göre, KTK'nın temel eşitliği;

$$X=T + E$$

şeklindedir (Güler, Uyanık, Teker, 2012:2).

Ölçme sürecinde ölçmeye karışan hatalar yüzünden gerçek değer doğrudan elde edilemez. Ölçmede amaç gerçek değer bulunmak istenmesi olsa da, gerçek değer doğrudan elde edilemediği için, gözlenen ölçme sonuçları yardımıyla kestirilmeye çalışılır. (Baykul, 2000). Baykul (2000)'e göre KTK'nın temel özellikleri:

1) Tekrarlı ölçmelerde, rastgele hata puanları negatif ve/veya pozitif yönde değişeceğinden, hataların evrendeki dağılımının ortalaması sıfıra eşit olur.

- 2) Gerçek puanlar ile hata puanlarının korelasyonu sıfırdır.
- 3) Aynı hata puanları arasındaki korelasyon sıfırdır.

Klasik test kuramında, gerçek puan varyansı ve hata varyansının toplamından gözlenen puan varyansı oluşmaktadır. Gerçek puan varyansının gözlenen puan varyansına oranı güvenilirlik katsayısını vermektedir (Lord & Novick, 1968). Bu katsayı '0' (güvenilir değil) ile '1' (mükemmel güvenilirlik) arasında değişen değerler almaktadır.

$$\sigma_{\text{gözlenen}}^2 = \sigma_{\text{gerçek}}^2 + \sigma_{\text{hata}}^2$$

$$\rho = \frac{\sigma_{\text{gerçek}}^2}{\sigma_{\text{gözlenen}}^2}$$

Gerçek varyans hariç kalan tüm varyansın farklı hata kaynaklarından gelebileceği düşünülmektedir. (Güler, 2008). Klasik test kuramında, güvenilirliğin anlamına ve ele alınan hata kaynağına göre güvenilirlik hesaplama yöntemleri farklılıklar göstermektedir (Baykul, 2000). Örneğin, güvenilirlik test-tekrar-test yöntemi ile (bir testin iki ya da daha fazla sayıdaki uygulaması) elde edilen puanlar arasındaki ilişki olarak tanımlanmaktadır. Burada güvenilirlik 'kararlılık' veya 'tutarlılık' anlamında kullanılmaktadır. Uygulamalar arasında geçen süre (zaman) ise test-tekrar-test yönteminde hata kaynağı olarak düşünülmektedir. Ölçme aracında bulunan maddelerden elde edilen puanların tutarlılığına 'iç tutarlılık' anlamında güvenilirlik denilmektedir. Ölçme aracındaki maddeler ise hata kaynağı olarak kabul edilmektedir. Eşdeğer (paralel) formlar yönteminde ise güvenilirlik iki paralel formun uygulanması sonucu elde edilen puanların 'tutarlılığı' şeklinde tanımlanmaktadır. Buradaki hata kaynağı ise ölçme aracı olarak değerlendirilmektedir. Birden fazla puanlayıcının yer aldığı ölçmelerde puanlayıcıların ölçülen aynı özelliğe verdikleri puanlar arasındaki tutarlılığa da 'puanlayıcılar arası güvenilirlik' denilmektedir ve puanlayıcılar hata kaynağı olarak düşünülmektedir (Güler, 2008).

Genel olarak bakıldığında, Hamilton & Swaminathan (1985) Klasik Test Kuramının sınırlılıklarını şu şekilde sıralamıştır.

- 1) Test geliştirmede kullanılan madde indekslerinden olan madde güçlük ve madde ayırıcılık indeksleri testin uygulandığı cevaplayıcı grubunun yeteneğine bağlı olmaktadır.



- 2) Başarı ve davranış testlerinin çoğu orta yetenek düzeyindeki öğrenciler için hazırlanmıştır. Yüksek ya da düşük yetenek düzeyindeki öğrenciler için kesin tahminler sağlamamaktadır.
- 3) Klasik Test Kuramının en önemli kavramlarından olan güvenilirlik, paralel formlarla tanımlanmaktadır. Paralel ölçümler pratikte oldukça zordur. Çünkü bireyler testin ikinci uygulamasında asla aynı kalamamaktadır.
- 4) Cevaplayıcıların, test maddeleriyle karşılaştıklarında nasıl performans gösterdikleri KTK ile karşılaştırılmamaktadır.
- 5) Ölçme hata varyansları tüm cevaplayıcılar için aynı kabul edilmektedir.
- 6) KTK ve uygulamaları; test geliştirme, test puanlarının eşitlenmesi, yanlı maddelerin belirlenmesi, ve maddelerin maksimum ayırıcılık gücü hakkında yeterli bilgi vermemektedir.

#### **1.8.1.1. Puanlayıcılar-arası Güvenirlik**

Değerlendirme yapılırken, elde edilen sonuçların puanlayıcıdan puanlayıcıya farklılık göstermesindenense, benzer çıkması beklenmektedir. Örneğin, ehliyet almak için direksiyon sınavına giren bir birey, sınav sonucunun kendisini değerlendiren kişiye göre değil, kendi performansına göre sonuçlandığına inanmaktadır. Maalesef, puanlayıcının olduğu değerlendirmelerde sonuç her zaman bu şekilde olmamaktadır (Cohen ve Swerdlik, 1996). Ölçme sonuçları, tesadüfi hatalardan olduğu kadar, puanlayıcıdan kaynaklanan nedenlerden (örn: tecrübe, farkındalık vb.) de etkilenmektedir. Puanlayıcılar aynı kurallar çerçevesinde değerlendirme yapsalar da genellikle farklı sonuçlar elde edilmektedir (Lorber, 2006).

Puanlayıcılar arası güvenilirlik, iki ya da daha fazla puanlayıcı arasındaki uyum ve tutarlılığın derecesidir. Puanlayıcılar arası güvenilirliği belirlemek için çeşitli yöntem ve yaklaşımlar bulunmaktadır. Bu yöntemlerden bazıları; uyum yüzdesi, korelasyon teknikleri, korelasyon katsayısıyla birlikte ortalamaların karşılaştırılması (t testi veya varyans analizi), Cohen'in kappası ( $\kappa$ ), ağırlıklandırılmış kappa ( $\kappa_w$ ), Fleiss'in kappası ve Kendall'in uyum katsayısı (W) dir. Bu tekniklerden hangisinin kullanılacağı, araştırmacının soyutlamak istediği ölçme hatasına, ölçek türüne (sınıflama, sıralama, eşit aralıklı, eşit oranlı) vs. göre çeşitlilik göstermektedir (Goodwin, 2001).

*Korelasyon teknikleri:* İki deęişken daęılımı arasındaki iliřkinin hesaplanmasında birçok korelasyon teknięi kullanılabilir. Ancak bu teknikler arasında en çok bilinen ve kullanılan korelasyon teknięi, eřit aralıklı ve oranlı ölçeklerdeki deęişkenlere uygulanabilen Pearson çarpım moment korelasyon katsayısıdır (Yelboęa, 2008). Pearson korelasyon katsayısı -1.00 ile +1.00 arasında deęerler alır. Korelasyon katsayısının +1.00 olması, iki puanlayıcının puanları arasında mükemmel pozitif bir iliřki; -1.00 olması, mükemmel negatif bir iliřki olduęunu; 0.00 olması ise iliřki olmadıęını ifade eder. Bu katsayının, mutlak deęer olarak, 0.00-0.30 arasında bir deęer alması, düşük; 0.30-0.70 arasında bir deęer alması, orta; 0.70-1.00 arasında bir deęer alması ise yüksek; düzeyde bir iliřki olarak tanımlanabilir (Büyüköztürk, 2002).

Çokça bilinmesi ve kolay yorumlanabilmesi bakımından, Pearson çarpım moment korelasyon katsayısı oldukça avantajlıdır. Öte yandan, Hartmann (1997) tarafından dezavantajları řu řekilde sıralanmaktadır:

- 1) Grubun büyüklüęünden doğrudan etkilenmesi
- 2) İki deęişken arasındaki doğrusal iliřkiyi hesaplaması
- 3) Ortalamadan bağımsız oluşu

*Korelasyon katsayısıyla birlikte ortalamaların karşılaştırılması:* Pearson çarpım moment korelasyonu, ortalamadan bağımsız olması sebebiyle, puanlayıcıların verdikleri puanlar arasındaki benzerlik ve farklılıklar hakkında bilgi verememektedir. Örneęin puanlayıcılar arasında puanlamada katılık/cömertlik bakımından ciddi bir fark olmasına rağmen, Pearson çarpım moment korelasyon katsayısı pozitif ve bire yakın çıkabilmektedir. Ancak bu durum, puanlayıcı puanları arasında farklılık olmadıęının deęil, sadece puanların birlikte deęişim gösterdięinin kanıtıdır. Daha doğru kestirimler yapabilmek için, Pearson çarpım moment korelasyon katsayısı ile birlikte ortalamaların karşılaştırılması da gerekmektedir. Goodwin (2001)'e göre, puanlayıcılar arası güvenilirlięin belirlenmesinde korelasyon katsayısı ile birlikte puanlayıcılardan elde edilen puan ortalamalarının, eřitirilmiş gruplar için t testi ya da tekrarlı ölçmeler için ANOVA kullanılarak karşılaştırılması daha faydalı olacaktır.

*Uyum yüzdesi:* Uyum yüzdesi, puanlayıcılar tarafından aynı puanı alan öğrenci sayısının, ölçmeye katılan tüm öğrencilerin sayısına bölünmesiyle yapılmaktadır

(alıntılayan Güler, 2008); (aktaran Meyer, 1999). Kolay anlaşılır olmasının yanında, korelasyon tekniklerinin aksine tüm ölçek türlerine (sınıflama, sıralama, eşit aralıklı, eşit oranlı) uygulanabilir olması en büyük avantajıdır. Ayrıca, puanlayıcılar arası uyumsuzluğun nerelerden kaynaklandığı hakkında bilgi vermesi nedeniyle, puanlayıcı eğitimlerinde ve puanlayıcılara verilen eğitimin yetersiz kaldığı durumlarda fayda sağlamaktadır. En büyük dezavantajı ise tesadüfen ortaya çıkan uyumu hesaplayamamasıdır (Goodwin, 2001).

*Cohen'in kappası:* Basit uyum yüzdesinin, şansla ortaya çıkan uyumu hesaplayamaması nedeniyle; alternatif bir puanlayıcılar arası güvenilirlik tekniği olarak Cohen'in kappası, Cohen tarafından 1960 yılında tasarlanmıştır. Eğer puanlayıcılar, şansa bağlı olarak aynı seviyede uyum gösterirlerse, kappa sıfıra eşit çıkar. Eğer uyum, şansa ortaya çıkabilecek uyum düzeyinin üzerinde çıkarsa, kappa sıfırdan daha büyük bir değer alacaktır. Kappa eşitliği;

$$K = \frac{P_0 - P_C}{1 - P_C}$$

denklemleri ile ifade edilir. Bu denklemde,  $P_0$  gözlenen uyum oranını;  $P_C$  şansla ortaya çıkabilecek uyum oranını ifade etmektedir. Kappa istatistiği, +1.0 ile -1.0 arasında değerler almaktadır. Ancak; negatif değerler, şansa ortaya çıkabilecek uyumun altında uyum düzeyini ifade edeceği için göz ardı edilmektedir. Bu istatistiğin en büyük dezavantajı, tüm uyumsuzlukları aynı seviyede göz önünde bulundurmaktadır (Goodwin, 2001).

Landis ve Koch (1977) tarafından; kappa değerinin, 0.20 ve altında olması, zayıf uyum; 0.21-0.40 arası olması, ortanın altında uyum; 0.61-0.80 arasında olması, iyi düzeyde uyum; 0.81-1.00 aralığında olması ise çok iyi düzeyde uyum olarak tanımlanmaktadır.

*Ağırlıklandırılmış kappa:* Kappa istatistiğinin, tüm uyumsuzlukları eşit düzeyde ele alması sebebiyle Cohen, ağırlıklandırılmış kappa istatistiğini geliştirmiştir. Ağırlıklandırılmış kappa istatistiğinde, bazı uyumsuzluklara hiç ağırlık vermezken, bazı uyumsuzluklara kısmi ağırlıklandırma uygulanmaktadır. Ağırlıklandırılmış kappa istatistiğinin dezavantajları şu şekilde sıralanmaktadır:

- 1) İki puanlayıcının puanlarının marjinal (uçlardaki) dağılımları aynı olduğu zaman, bu katsayı teorik olarak en yüksek değere yani 1.00 eşit olabilmektedir.
- 2) Az ya da daha çok madde ya da puanlayıcı olması durumlarında, güvenilirlik hakkında kestirimlerde bulunmak için kullanılan Spearman-Brown formülü gibi bir formül, kappa istatistiklerinde kullanılamamaktadır.
- 3) Puanlayıcılar dışında başka potansiyel hata kaynaklarının olması durumunda, bu kaynaklar güvenilirlik hesaplanmasında ele alınamamaktadır (Goodwin, 2001).

*Fleiss'in kappası:* Cohen'in kappası, bütün uyumsuzluklar eşit olarak ele alınacağı zaman, ağırlıklandırılmış kappa ise sadece belirlenmiş bazı uyumsuzluklara ağırlık verileceği zaman faydalı tekniklerdir. Ancak ikiside, sadece 2 puanlayıcının olduğu durumlarda kullanılabilir. Bunlardan farklı olarak, Fleiss'in kappası, tüm puanlayıcıların rastgele uyumlarının üzerinde gözlenen uyumlarının derecesini verir. Fleiss'in kappa eşitliği

$$\kappa = \frac{P - P_0}{1 - P_0} \text{ ile bulunur.}$$

$1 - P_0$ , şansla ortaya çıkabilecek uyumun büyüklüğünü ifade ederken;  $P - P_0$ , gözlenen uyumun büyüklüğünü vermektedir. Bu katsayı, 0 ile 1 arasında değerler almaktadır. 1 değeri mükemmel uyum anlamına gelmektedir (Fleiss, 1971).

*Kendall'in uyum katsayısı:* İki'den fazla puanlayıcının olduğu sıralı ölçeklerde, puanlayıcılar arası güvenirliliği belirlemek için kullanılan parametrik olmayan bir istatistiksel tekniktir. İlk olarak x tane gözlem, y tane özelliğin her birinde elde edilen puanlara göre sıralanabileceği gibi, y adet karar verici x tane gözlemi aynı özelliğe göre sıralayabilir. Bu gibi

durumlarda kullanılan Kendall'in uyum katsayılarının varsayımları aşağıdaki gibidir (Kıroğlu, 2001).

- 1) Veriler x gözlem ya da birey üzerinde y adet ölçmeden oluşur.
- 2) Ölçme ölçeği en azından sıralıdır.
- 3) Kaydedilen gözlemler sıra sayılarından oluşabilir. Orijinal veriler sıralı değilse sıralanmalıdır.

Kendall'ın uyum katsayısının eşitliği aşağıdaki gibidir (Howell, 2012):

$$W = \frac{12 \sum T_j^2}{K^2 N (N^2 - 1)} - \frac{3(N+1)}{N-1}$$

Bu eşitlikte:

K: puanlayıcıların sayısı

N: Puanlanan görev ya da madde sayısı

$T_j$ : Her bir madde ya da göreve tüm puanlayıcıların verdiği puanların toplamını göstermektedir.

Kendall'ın uyum katsayısına ilişki yokluk hipoteziyle de yorum yapılabilmektedir. 7 ve daha fazla puanlayıcının olduğu durumlarda  $X^2$  ile hesaplama yapılır.

$X^2_{(N-1)}$  = k (N-1)W değeri N-1 serbestlik derecesinde yaklaşık  $X^2$  dağılımını göstermektedir (Howell, 2002).

### 1.8.2. Genellenebilirlik Kuramı

Daha önceden çeşitli makalelerde tartışılmış olsa da, Genellenebilirlik kuramının en etkili sunumu 1972 yılında Cronbach, Gleser, Nanda ve Rajaratman tarafından gerçekleştirilmiştir. Daha sonra, Brennan (2000a, 2000b, 2001), Hoyt ve Melby (1999), Kieffer (1998), Marcoulides (1999, 2000), Strube (2000), Shavelson, Webb, ve Rowley (1989), ve Webb ve diğerleri (1988) tarafından yapılan katkı ve görüşlerle genellenebilirlik kuramı günümüzdeki halini almıştır (Mushquash ve O'connor, 2006).

Genellenebilirlik kuramı, davranış değerlendirilmesinde güvenilirliğin hesaplanmasını, güvenilir gözlemlerin G ve K çalışmalarıyla tasarlanıp, araştırılmasını ve çeşitli hata kaynaklarını göz önünde bulundurarak tek bir güvenilirlik katsayısının bulunmasını sağlayan istatistiksel bir kuramdır. Genellenebilirlik kuramının amacı, ölçme sonuçlarını farklı varyans kaynaklarına ayırıp, yorumlayıp, tanımlayarak ölçme konusu olan bireyler veya nesnelerin gözlenen puanlarının Klasik test kuramındaki gerçek puan kavramına karşılık gelen evren puanlarına genellemesini sağlamaktır (Atılğan, 2004). Yani, genellenebilirlik kuramı sadece bir ölçme aracının farklı durumlarda ne kadar güvenilir olduğu ile ilgilenmez, aynı zamanda elde edilen ölçme sonuçlarının evrene ne kadar genellenebilir olduğunu araştırır (Eason, 1989).

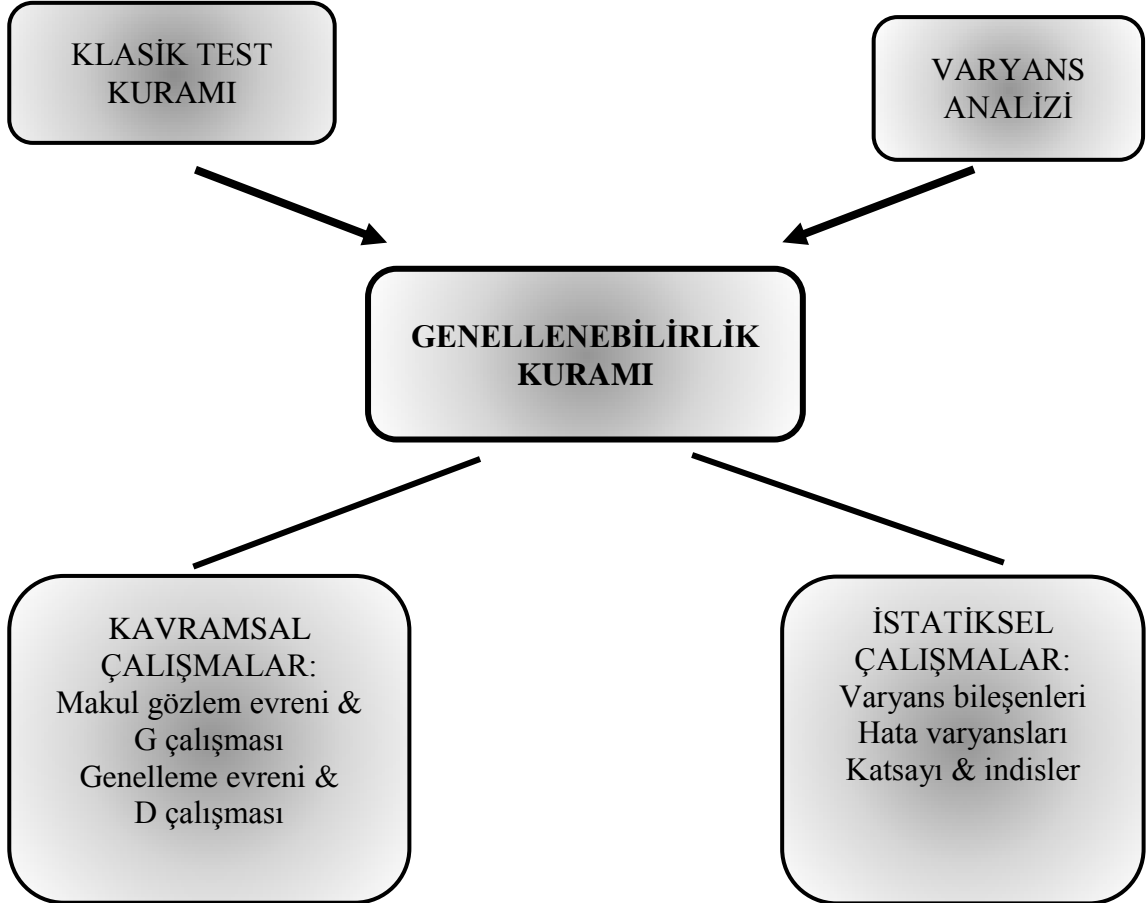
Shavelson ve Webb (1991)'e göre G kuramı, 4 farklı açıdan klasik test kuramının genişletilmiş halidir:

- 1) Genellenebilirlik kuramı, çoklu varyans kaynaklarını tek bir analizde ele alır.
- 2) Varyans kaynaklarının büyüklüklerinin ayrı ayrı belirlenmesini sağlar.
- 3) Güvenirlik için, bağıl kararlarda G katsayısının, mutlak kararlarda phi katsayısının hesaplanmasına olanak sağlar.
- 4) Belirli bir amaca bağlı olarak, ölçme hatasının en aza indirgenebileceği ölçmelerin düzenlenmesine imkân tanır.

Genellenebilirlik kuramının temeli, istatistiksel olarak varyans analizine (ANOVA) dayanmaktadır. Araştırmacılar önemli gördükleri bağımsız değişkenlerin etkilerini belirlemeye ve tahmin etmeye çalışırken, ölçme uzmanları G kuramını kullanarak değişkenlik kaynaklarının miktarlarını belirlemeye ve kestirmeye çalışırlar. Ölçme uzmanlarının belirlemeye çalıştığı bu değişkenlik, evren puanlarından ve çeşitli hata kaynaklarından gelen varyanstır (Güler, Uyanık, Teker, 2012).

Klasik test kuramında farklı güvenilirlik hesaplanma yöntemleri bulunmaktadır. Her yöntemde güvenilirlik, anlamına ve ele alınan tek hata kaynağına göre farklı adlar almaktadır (Baykul, 2000). Güvenirlik hesaplanırken, test-tekrar-test yönteminde ele alınan olası hata kaynağı zaman, paralel (eşdeğer) formlar yönteminde formlar, iç tutarlılık anlamında hesaplama yapılırken ise hata kaynağı olarak görevler ya da maddeler ele alınır (Brennan, 2001; Shavelen ve Webb, 1991; Güler ve Gelbal, 2010). Bu da, klasik test kuramının en önemli sınırlılıklarındandır. Örneğin; bir uyuşturucu rehabilitasyonunda, 4 uzman tarafından 2 farklı seansta danışanların tutumları değerlendiriliyor. Bu değerlendirmede, puanlayıcılar (uzmanlar) arası varyansın ve seanslar arası varyansın hangi ölçüde hata varyansına etki edeceğinin belirlenmesi faydalı olacaktır. Böyle bir durumda, tanımlanmamış tek bir hata bileşeni veren klasik test kuramının uygulanması doğru olmayacaktır (Crocker ve Algina, 2006). Genellenebilirlik kuramı ya da G kuramı, birden fazla hata kaynağının olduğu durumlarda güvenilirlik kestirimi için geliştirilen bir yaklaşımdır. Tek bir güvenilirlik katsayısı vermekle kalmaz, aynı zamanda göz önünde bulundurulduğu takdirde gelecekteki uygulamalara katkı sağlayacak hata varyansları hakkında da bilgi verir (Cardinet, Johnson ve Pini, 2009).

Klasik test kuramı ve ANOVA, genellenebilirlik kuramının ebeveynleri gibi düşünülse de; çocuk, ebeveynlerinin birleşiminden hem daha basit hem de daha karmaşıktır. Genellenebilirlik kuramını anlamak, onun kökenini anlamaktan daha fazlasını gerektirir. (Brennan, 2001).



**Şekil 1.1. Genellenebilirlik Kuramının Kökeni ve Kavramsal Çerçevesi (Brennan, 2001)**

Genellenebilirlik kuramı, klasik test kuramının bir uzantısı olarak kabul edilir (Brennan, 2001; Shavelson ve Webb, 1991; MacMillian, 2000). Ancak, iki kuram arasında önemli farklılıklar da bulunmaktadır (Shavelson ve Webb, 1991).

- 1) GK ve KTK hata kaynaklarını ele alış bakımından birbirinden ayrılır. GK, ölçmeye karışan hataların birden fazla kaynaktan geldiğini kabul ederken, KTK bu hata kaynaklarının hepsinin tek bir hata değişkeninden geldiğini kabul eder.
- 2) GK, farklı kaynaklardan gelen hataların (puanlayıcı, madde, zaman vb.) hepsini birlikte göz önünde bulundururken aynı zamanda bu hataların

birlikte etkileşimini de hesaplar. Bu yönüyle GK, farklı kaynaklarından gelen hata varyanslarını dikkate alarak ölçümlerin güvenilirliğini maksimuma çıkartabilir (Eason, 1989).

- 3) KTK güvenilirlik hesaplamasında sadece bağıl değerlendirmeler için güvenilirlik hesaplarken, GK hem bağıl değerlendirme hem de mutlak değerlendirmeler için güvenilirlik katsayıları üretir.
- 4) Genellenebilirlik kuramında, daha yüksek güvenilirlik elde edebilmek için alternatif K-çalışmaları yapılarak değişkenlik kaynaklarının koşullarının sayısı (madde, puanlayıcı, zaman vs.) belirlenebilir. Oysa klasik test kuramında sadece Spearman-Brown formülü kullanılarak bir değişkenlik kaynağının (madde) sayısı hesaplanabilir.
- 5) GK, genel anlamda geçerlik ve güvenilirlik arasındaki farkı ortadan kaldırır. KTK'da yapı geçerliği, ölçülmek istenen bir yapının, bir diğer deyişle, bireylerde var olduğu kabul edilen bir özelliğin, ölçme sonucunda ortaya konulabilme derecesi olarak yorumlanabilir (Baykul, 2000). Genellenebilirlik kuramının amacı, ölçme konusu olan bireyler ya da nesnelere gözlenen puanlarının evren puanlarına (KTK'daki gerçek puan kavramına karşılık gelmektedir) doğrulukla genellenmesini sağlamaktır (Atılğan ve Tezbaşaran, 2005). Bu, aynı zamanda yapı geçerliğinin de bir kanıtı olarak görülmektedir.

Tüm bunlar göz önüne alındığında özellikle birden fazla değişkenlik kaynağının etkin olduğu çok sayıda puanlayıcının kullanıldığı ya da birden fazla kez ölçmenin yapıldığı durumlarda GK, KTK'ya göre daha avantajlı bir kuram olarak karşımıza çıkmaktadır (Güler, 2011).

G kuramı açıklanırken, konunun daha iyi anlaşılması için bazı kavram ve terminolojilerin açıklanması önem arz etmektedir. Genellenebilirlik kuramında, bir ölçmedeki zaman, puanlayıcı, madde, formlar vb. gibi potansiyel hata kaynaklarına, yüzey (facet, değişkenlik kaynağı) adı verilmektedir. Bir değişkenlik kaynağının her bir düzeyi koşul (condition) olarak adlandırılmaktadır (Brennan, 2001; Shavelson ve Webb, 1991; Crocker ve Algina, 1986). Buradaki yüzey (facet) ifadesi varyans analizindeki faktörler gibi düşünülebilir. Koşullar da varyans analizindeki faktörlerin düzeyleri gibi ele alınabilir (Rentz, 1987). Başka bir örnekle



açıklarsak, bir ölçmede formlar değişkenlik kaynağı ise, form sayısı bu değişkenlik kaynağının koşulu olur.

Ölçme sonucunda, istenilen kararların alınacağı ölçmenin hedefine ölçme objesi (object of measurement) denir. Ölçme objesi genellikle sistematik varyans içeren bireylerdir. Çünkü bireyler, doğası gereği farklılık gösterir. Fakat bu istenilen bir durumdur. Ölçme objesinin içerdiği sistematik varyans, hata varyansı olarak kabul edilmez. Bu yüzden de genellenebilirlik kuramında yüzey olarak düşünülmez. Ancak bireyler her zaman ölçmenin objesi olmak zorunda değildir. Ölçme durumuna göre, sistematik varyans içeren her şey (madde, durum vb.) ölçme objesi olabilir (Eason, 1989, Mushquash ve O'connor, 2006).

Araştırmacının, genellemek istediği yüzeyin koşullarına genelleme evreni (universe of generalization), alınabilecek tüm koşulların oluşturacağı evrene ise kabul edilebilir gözlemler evreni (universe of admissible observation) denir (Crocker ve Algina, 1986). Kabul edilebilir gözlemler evreninden alınan puanların ortalaması ise evren puanını (universe score) verir. Bu da klasik test kuramındaki kişinin gerçek puanı tanımlamasına benzer (Brennan, 2010; Shavelson ve Webb, 1991; Keiffer, 1998). Hata puanları varyansı ise, G kuramı içinde tanımlanan, farklı hata kaynakları ve bunların etkileşiminden gelen hata bileşenleri ile kalan hata varyansından oluşur (Atılğan, 2004).

#### **1.8.2.1. Genellenebilirlik (G) Çalışması & Karar (K) Çalışması**

Genellenebilirlik kuramında güvenilirlik G çalışması (Generalizability study) ve K çalışması (Decision (D) study) olarak ifade edilen iki tür çalışma yer almaktadır.

G çalışmasının amacı kabul edilebilir gözlemler evrenine ilişkin varyans bileşenlerini belirlemek ve bu varyans kaynakları hakkında olabildiğince bilgi edinmektir. Bu aşamada, ölçme objesi ve yüzeylerden oluşan gerçek veriler üzerinde çalışmalar yapılmaktadır (Brennan, 2001; Shavelson ve Webb, 1991; Keiffer, 1998). Varyans bileşenleri belirlendikten sonra 'Eğer ki?' sorularına cevap aramak için K çalışmasına geçilir. Bu aşama sürecinde araştırmacı, madde sayıları, form sayıları gibi koşullar üzerinde oynamalar yaparak daha yüksek güvenilirlik ve daha düşük hata içeren sonuçlar elde etmeye çalışır (Kieffer, 1998). Örneğin, "Daha fazla puanlayıcı kullandığımda ne olur?", "Görev sayısını azaltırsam ölçme aracımın güvenilirliğinde kaybım çok fazla olur mu?" vb.

sorularının cevabı K çalışması yapılarak araştırılır. Sonuç olarak, tek bir G çalışmasından elde edilen aynı varyans kestirimlerine dayalı pek çok K çalışması düzenlenebilir (Güler, 2008). Klasik test kuramında Spearman-Brown formülüyle madde sayısında değişiklikler yapılarak güvenilirlik kestirimi yapılabilmektedir. K çalışmasında ise güvenilirliği artırabilmek için sadece madde sayısında değil, tüm yüzeylerin sayısında değişiklikler yapılarak güvenilirlik kestirimleri elde edilebilmektedir. Böylece K çalışmaları en yüksek güvenilirliğin ve en etkili ölçme durumlarının kestirilmesine yardımcı olabilmektedir (Lee ve Fitzpatrick, 2003).

Unutulmamalıdır ki, K çalışmasında sadece G çalışmasında kullanılan yüzeyler kullanılabilir. Örneğin, madde ve puanlayıcıdan oluşan 2 yüzeyli bir G çalışması sonucunda yapılan K çalışmasında, zaman yüzeyinin etkisini test edemezsiniz. Ayrıca, G çalışmasında kullanılan her yüzeyin K çalışmasında kullanılmasına gerek yoktur (Kieffer, 1998).

#### **1.8.2.2. Tesadüfi Değişkenlik Kaynakları & Sabit Değişkenlik kaynakları**

Genellenebilirlik kuramında tesadüfi ve sabit olmak üzere iki tür değişkenlik kaynağı (yüzey) vardır. Hangi değişkenlik kaynağının kullanılacağı araştırmacının kararına bağlıdır. Eğer araştırmacı örneklemin ötesinde genelleme evrenine genelleme yapmak istiyorsa, bu durumda değişkenlik kaynağı tesadüfi olarak ele alınacaktır. Öte yandan, eğer araştırmacı örneklemin ötesinde genellemeye gitmek istemiyorsa, değişkenlik kaynağı sabit olarak ele alınacaktır. (alıntılayan Alharby, 2006); (aktaran Öztürk, 2011).

Örneklemin evrenden küçük olduğu ve örneklemin evrenden tesadüfi olarak çekildiği ya da değiştirilebilir olduğu durumlarda değişkenlik kaynağı tesadüfi olarak kabul edilmektedir. Değişkenlik kaynağı tesadüfi olarak davranıyorsa, “Bu örneklemdaki koşulları evrende aynı boyuttaki başka bir koşul grubuyla değiştirmek ister miyim?” sorusu sorulmalıdır. Örneğin, araştırmacı fen başarı testinde bulunan 40 maddeyi, başka bir 40 madde örneklemiyle değiştirmek istediğinde, değişkenlik kaynağı tesadüfi kabul edilir (Shavelson ve Webb, 1991).

Sabit değişkenlik kaynağı ise varyans analizindeki sabit faktör gibi düşünülebilmektedir. Araştırmacı tarafından kasıtlı olarak seçilir ve evrene genellenmesiyle ilgilenilmemektedir. Sabit değişkenlik kaynağı, üzerinde çalışılan evrenin küçük olması nedeniyle sadece evren üzerinde çalışılması gereken

durumlarda kullanılmaktadır ( Crocker ve Algina, 1986; Brennan, 2001; Shavelson ve Webb, 1991).

Değişkenlik kaynağının tesadüfi ya da sabit olması ölçmenin genellenebilirliğini etkilemektedir. Unutulmamalıdır ki, bir ya da daha fazla değişkenlik kaynağını sabitlemek, hata varyansını düşürmekte ve güvenilirlik katsayılarında artışa neden olmaktadır. Ancak bu durum, yapılacak yorumların kısıtlanmasını da beraberinde getirmektedir (Brennan, 2001).

### **1.8.2.3. Çaprazlanmış (Crossed) Desen & Yuvalanmış (Nested) Desen**

Genellenebilirlik kuramında evren sahip olduğu değişkenlik kaynaklarına göre üç başlık altında incelenmektedir. Bunlardan ilki, bir potansiyel değişkenlik kaynağına sahip olan tek yüzeyli evrenlerdir (Eser ve Gelbal, 2011). Örneğin; 10 öğrenci, '0' ve '1' olarak puanlanan ve 4 seçenekli 10 maddeden oluşan bir matematik başarı testini cevaplasın. Burada öğrenci (birey) ölçmenin objesi, ölçmede yer alan tek yüzeyde madde yüzeyidir. İkincisi, 2 tane değişkenlik kaynağı bulunan iki yüzeyli evrenlerdir. Örneğin, 10 öğrenci, 10 maddelik testi cevaplasın ve 2 puanlayıcı da cevapları puanlasın. Burada öğrenci (birey) ölçme objesi iken, ölçmede yer alan 2 yüzey de madde yüzeyi ve puanlayıcı yüzeyidir. Sonuncusu da üç ya da daha çok yüzeyli evrenlerdir.

Değişkenlik kaynaklarının ele alınış şekillerine göre G kuramında 2 tür desen vardır. Bir ölçmedeki değişkenlik kaynaklarının (yüzeylerin) tüm koşulları, diğer değişkenlik kaynaklarının tüm koşullarını etkiliyorsa bu desene çaprazlanmış desen denir. Bu desende yüzeyler arasına 'x' işareti konulur. Bir ölçmedeki bir değişkenlik kaynağının bazı koşulları, diğer bir değişkenlik kaynağının bazı koşullarını etkiliyorsa bu desene de yuvalanmış desen denir. Bu desen de yüzeyler arasına ':' işareti konularak gösterilir (Brennan, 2001; Shavelson ve Webb, 1991; Mushquash ve O'connor, 2006).

İki yüzeyli evrenlerde bu desenlere örnek vermek gerekirse, eğer her bir öğrenci (b) her bir maddeyi (m) cevaplandırır ve her bir puanlayıcı (p) da her bir öğrencinin tüm maddelerini puanlarsa bu desen tümüyle çaprazlanmış bir desen olup, "b x m x p" şeklinde gösterilir. Eğer farklı öğrenciler (b) farklı maddeleri (m) cevaplarsa ve farklı puanlayıcılar (p) puanlamayı yaparsa b desene de tümüyle yuvalanmış desen denir ve b : m : p olarak gösterilir (Güler, Uyanık, Teker, 2012).

Çalışmalarda genellikle tümüyle çaprazlanmış desen tercih edilir. Çünkü bu desen, tüm değişkenlik kaynaklarından ve bunların etkileşiminden meydana gelen hataların kestirimine olanak sağlar (Mushquash ve O'connor, 2006).

Araştırma da kullanılan desenlerden biri iki yüzeyli tümüyle çaprazlanmış  $b \times m \times p$  tesadüfi desendir. Burada maddeler ve puanlayıcılar değişkenlik kaynağı olup, bireyler ise ölçme objesidir. Bu durumda, toplam gözlenen varyans aşağıda ifade edilen yedi bağımsız varyans bileşeninden oluşur (Alharby, 2006; akt. Güler, 2008):

$$\sigma^2 (X_{bmp}) = \sigma^2_b + \sigma^2_m + \sigma^2_p + \sigma^2_{bm} + \sigma^2_{bp} + \sigma^2_{mp} + \sigma^2_{bmp,e}$$

Her bir varyans bileşeni, Tablo X de gösterildiği gibi varyans analiziyle kestirilebilir:

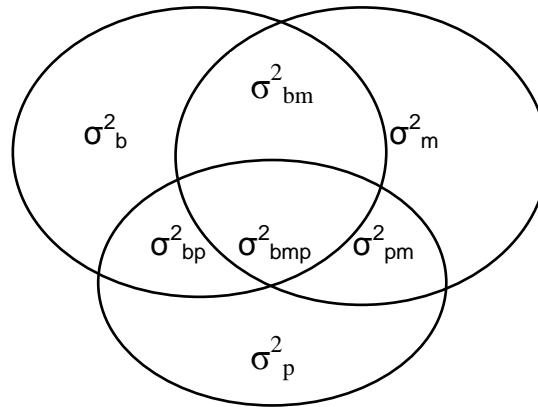
**Tablo 1.1. İki Yüzeyli Tesadüfi Desen için Varyans Bileşenlerinin Kestirilmesi**

<i>Varyans Kaynağı</i>	<i>Kareler Toplamı</i>	<i>Sd</i>	<i>Kareler Ortalaması</i>	<i>Kestirilen Varyans Bileşenleri</i>
<i>Birey (b)</i>	$SS_b$	$n_b-1$	$MS_b=SS_b/n_b-1$	$\sigma^2_b$
<i>Puanlayıcı (p)</i>	$SS_p$	$n_p-1$	$MS_p=SS_p/n_p-1$	$\sigma^2_p$
<i>Madde (m)</i>	$SS_m$	$n_m-1$	$MS_m=SS_m/n_m-1$	$\sigma^2_m$
<i>b x p</i>	$SS_{bp}$	$(n_b-1)(n_p-1)$	$MS_{bp}=SS_{bp}/n_{bp}-1$	$\sigma^2_{bp}$
<i>b x m</i>	$SS_{bm}$	$(n_b-1)(n_m-1)$	$MS_{bm}=SS_{bm}/n_{bm}-1$	$\sigma^2_{bm}$
<i>p x m</i>	$SS_{pm}$	$(n_p-1)(n_m-1)$	$MS_{pm}=SS_{pm}/n_{pm}-1$	$\sigma^2_{pm}$
<i>b x m x p, e</i>	$SS_{bmp,e}$	$(n_b-1)(n_p-1)(n_m-1)$	$MS_{bmp,e}=SS_{bmp,e}/n_{bmp}-1$	$\sigma^2_{bmp}$

Bu tablodaki değişkenlik kaynakları şunlardır:

1. Birey (b)  $\sigma^2_b$  : Ölçme objesi olan bireylerin maddelerden aldıkları farklı puanlardan kaynaklanır. Bu değişkenlik kaynağı evren puanı değişkenliği olarak kabul edilir.
2. Madde (m)  $\sigma^2_m$  : Maddelerin her biri diğerlerine göre daha zor ya da daha kolaydır. Bu değişkenlik kaynağı, madde güçlükleri arasındaki farktan kaynaklanır.

3. Puanlayıcı (p)  $\sigma^2_p$  : Puanlayıcıların değerlendirmeleri arasındaki tutarsızlıktan kaynaklanır. Kimi puanlayıcılar diğerlerine göre daha cömert, kimileri daha katı puanlama yapabilirler.
4. Birey madde etkileşimi (b x m)  $\sigma^2_{bm}$  : Bazı bireyler bazı maddeleri cevaplarken kişisel sebeplerden ötürü (geçmiş yaşantılar vs.) daha avantajlı olabilirler. Bu değişkenlik kaynağı, birey madde etkileşiminin etkisinden kaynaklanmaktadır.
5. Birey puanlayıcı etkileşimi (b x p)  $\sigma^2_{bp}$  : Bazı puanlayıcılar bazı bireylere karşı daha katı ya da cömert olabilir. Bireylerin durumlarının bir puanlayıcıdan diğerine farklılık göstermesinden kaynaklanan değişkenlik kaynağıdır.
6. Madde puanlayıcı etkileşimi (m x p)  $\sigma^2_{pm}$  : Bazen bir puanlayıcı bir madde de çok cömertken, diğer bir madde de daha katı davranabilir. Puanlayıcıların bir maddeden diğerine tutarsızlıklarından kaynaklanan değişkenlik kaynağıdır.
7. Birey madde puanlayıcı etkileşimi (b x m x p)  $\sigma^2_{bmp,e}$  : Ölçme hatasından kaynaklanan artık değişkenlik kaynağıdır.



**Şekil 1.2. İki Yüzeyle Çaprazlanmış b x m x p Deseni İçin Varyans Bileşenleri**

#### **1.8.2.4. Bağlı (Görel) Hata Varyansı & Mutlak Hata Varyansı**

Davranış ölçmelerinde genellenebilirlik kuramı, bağlı ve mutlak değerlendirme olmak üzere iki tür değerlendirmeye olanak sağlamaktadır. Bağlı değerlendirmede, bireyin performansı içinde bulunduğu grubun performanslarına göre değerlendirilirken, mutlak değerlendirme de bireyin performansı, içinde bulunduğu

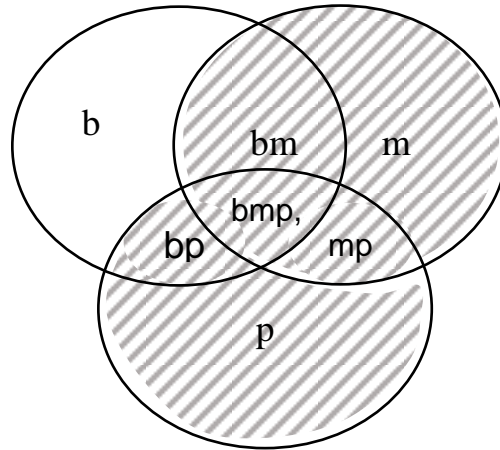
grubun performanslarına bakılmaksızın değerlendirilir (Volpe, McConaughy ve Hintze, 2009). Brennan (2001)'e göre, bağıl değerlendirmeler bireyin "diğerlerinden ne kadar iyi" olmasını göz önüne alırken, mutlak değerlendirmeler sadece bireyin "ne kadar iyi" olduğunu dikkate alır.

Bir başka deyişle, mutlak değerlendirme gruptan bağımsız olarak önceden belirlenmiş mutlak bir ölçüte göre yapılırken, bağıl değerlendirme gruba bağılı olarak önceden belirlenen bir norma göre yapılır. Bu nedenle GK'da ölçme hataları, bağılı ve mutlak kararlar için ayrı ayrı değerlendirilir (Yelboğa, 2008).

Mutlak hata varyansı bireylerin gözlenen ve evren puanları arasındaki farkın varyansı olarak tanımlanır ve  $\sigma^2(\Delta)$  şeklinde gösterilir. Tesadüfi tamamen çaprazlanmış iki yüzeyli desende mutlak hata varyansı aşağıdaki formülle hesaplanır:

$$\sigma^2(\Delta) = \frac{\sigma_m^2}{n'_m} + \frac{\sigma_p^2}{n'_p} + \frac{\sigma_{bm}^2}{n'_m} + \frac{\sigma_{bp}^2}{n'_p} + \frac{\sigma_{mp}^2}{n'_m n'_p} + \frac{\sigma_{bmp,e}^2}{n'_m n'_p}$$

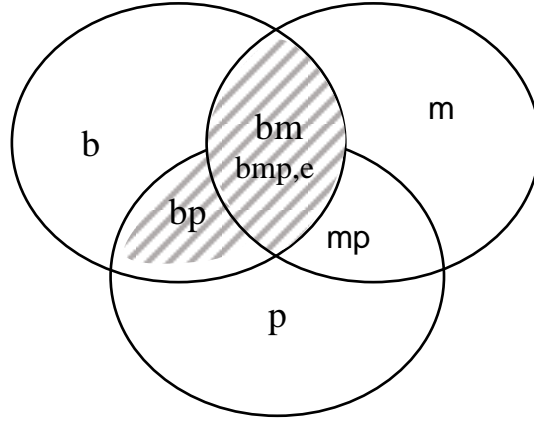
Mutlak kararlar için, evren puan varyansı hariç tüm kalan varyans kaynakları hataya katkıda bulunmaktadır.



**Şekil 1.3. Tesadüfi b x m x p Desende Mutlak Değerlendirmeler İçin Mutlak Hata Kaynağı**

Bağılı hata varyansı ise bireylerin gözlenen sapma puanları ve evren sapma puanları arasındaki farkın varyansı olarak tanımlanır ve  $\sigma^2(\delta)$  şeklinde gösterilir. Bağılı hata varyansı klasik test teoremindeki hata varyansına karşılık gelir. (Brennan, 2001). Tesadüfi tamamen çaprazlanmış iki yüzeyli desende bağılı hata varyansı aşağıdaki formülle hesaplanır:

$$\sigma^2(\delta) = \frac{\sigma_{bm}^2}{n'_m} + \frac{\sigma_{bp}^2}{n'_p} + \frac{\sigma_{bmp,e}^2}{n'_m n'_p}$$



**Şekil 1.4. Tesadüfi b x m x p Desende Bağlı Değerlendirmeler İçin Bağlı Hata Kaynağı**

#### **1.8.2.5. Genellenebilirlik Katsayısı (G-katsayısı) & Genellenebilirlik İndeksi (Phi-katsayısı)**

Genellenebilirlik kuramı, mutlak ve bağlı olmak üzere iki tür değerlendirmeye olanak sağlamasından ötürü bu değerlendirmeler için iki farklı güvenilirlik katsayısı bulunur. Bağlı değerlendirmeler için güvenilirlik katsayısı G-katsayısı, mutlak değerlendirmeler için ise güvenilirlik katsayısı Phi-katsayısı ( $\Phi$ ) hesaplanır (Brennan, 2001; Shavelson ve Webb, 1991; Goodwin, 2001; Crocker ve Algina, 1986).

Genellenebilirlik katsayısı (G), klasik test kuramındaki gerçek varyansın gözlenen varyansa oranı olan güvenilirlik katsayısına benzemektedir. Klasik test kuramındaki gerçek varyans, genellenebilirlik kuramında evren varyansı olarak adlandırılmaktadır. Bu sebeple, G katsayısı evren puan varyansının gözlenen puan varyansına oranı olarak hesaplanmaktadır. Gözlenen puan varyansı ise evren puan varyansı ( $\sigma^2$ ) ve bağlı hata varyansının ( $\sigma^2(\delta)$ ) toplanmasıyla elde edilir (Brennan, 2001; Shavelson ve Webb, 1991). İki yüzeyli tamamen çaprazlanmış b x m x p desende G ve  $\Phi$  katsayıları aşağıdaki formüllerle hesaplanır:

$$G \text{ katsayısı} = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_\delta^2}$$

Buradan  $\sigma_\delta^2$  yerine yazılırsa;

$$G \text{ katsayısı} = \frac{\sigma_b^2}{\sigma_b^2 + \frac{\sigma_{bm}^2}{n'_m} + \frac{\sigma_{bp}^2}{n'_p} + \frac{\sigma_{bmp,e}^2}{n'_m n'_p}}$$

bağıl değerlendirmeler için G katsayısı yukarıdaki eşitlikle bulunmuş olur.

G katsayısı hesaplanırken her bir ölçme objesinin, değişkenlik kaynağındaki diğer ölçme objelerinin puanlarının sıralaması arasındaki yerine bakılır, aldığı ham puanın ne kadar yüksek olduğuna bakılmaz. Bu sebeple elde edilen puanların puan sıralamasındaki yerinin önemli olduğu durumlarda bağıl G katsayısının kullanılması uygundur (Öztürk, 2011).

Mutlak değerlendirmelerde kullanılan  $\Phi$  katsayısı, evren puan varyansının ( $\sigma^2$ ), evren puan varyansı ve mutlak hata varyansı ( $\sigma^2(\Delta)$ ) nın toplamına oranı olarak hesaplanır.

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

Buradan  $\sigma_{\Delta}^2$  yerine yazılırsa;

$$\Phi \text{ katsayısı} = \frac{\sigma_b^2}{\sigma_b^2 + \frac{\sigma_m^2}{n'_m} + \frac{\sigma_p^2}{n'_p} + \frac{\sigma_{bm}^2}{n'_m} + \frac{\sigma_{bp}^2}{n'_p} + \frac{\sigma_{mp}^2}{n'_m n'_p} + \frac{\sigma_{bmp,e}^2}{n'_m n'_p}}$$

bağıl değerlendirmeler için G katsayısı yukarıdaki eşitlikle bulunmuş olur.

$\Phi$  katsayısı ise G katsayısına göre daha katı bir değer olup, hem ham puanların tutarlılığının derecesini hem de ölçmenin objesine ilişkin (genellikle bireylerdir) puanların sıralamasındaki tutarlılığın derecesini ortaya koyar. Bu sebeple,  $\Phi$  katsayısının belirli bir kesme puanının üzerindeki puanın önem arz ettiği performans ölçümlerinde kullanılması uygundur (Lee ve Frisbie, 1999).

Phi katsayısının paydası evren puan varyansı ve mutlak hata varyansının toplamından oluşmaktadır. G katsayısının paydası ise evren puan varyansı ve bağıl hata varyansından oluşmaktadır. Mutlak hata varyansı, bağıl hata varyansından, tanımları gereği, daha büyük bir değere sahiptir. Bu nedenle, Phi-katsayısı, G-katsayısından daha düşük bir değere sahip olur (Güler, Uyanık, Teker, 2012). İki güvenilirlik katsayısı da (G &  $\Phi$ ) 0 ve 1 arasında değerler alır. G ve  $\Phi$  katsayılarının yeterli ölçütleri isteğe bağlı olarak değişmektedir. Ancak bazı araştırmacılar G ve Phi katsayılarının 0,80'den büyük olması durumunda "yüksek"



olarak deęerlendirilebileceęini ifade etmektedir (Brennan, 2001). Shavelson ve Webb (1991) ise 0.80 ve üzeri genellenebilirlik katsayılarının anlamlı olduęunu söylemektedir.  $\Phi$  katsayısı, G katsayısına göre daha katıdır. G katsayısı özellikle tek yüzeyle tamamen çaprazlanmış desenlerde klasik test kuramındaki Cronbach alfa katsayısına benzemektedir (Sudweeks, Reeve, & Bradshaw, 2005; Güler, Güler, Eroęlu ve Akbaba, 2011).

## 2. İLGİLİ ARAŞTIRMALAR

Sub, Valiga ve Goa 1997 yılında Akademik Danışmanlık Anketi'ndeki (Academic Advising Survey, AAS) öğrenci puanlamalarının analizi genellikle Cronbach Alfa ile yapılması ve bu yöntemin güvenilirlik açısından çoğu zaman yetersiz kalması nedeniyle bu anketten elde edilen verileri genellenebilirlik kuramıyla çalışmışlardır. Veriler, lise sonrası eğitim veren 15 enstitüde bulunan 150 danışmanın her birinin 10 öğrenciye danışmanlık vererek, danışmaya katılan 1500 öğrencinin 18 maddelik Akademik Danışmanlık Anketi'ne verdikleri puanlar ile elde edilmiştir. Çalışmada (ö : e : d) x m (öğrenci, enstitü, danışman, madde) deseni çalışılmıştır. Araştırmada ayrıca karşılaştırma yapılabilmesi için klasik test kuramında Cronbach Alfa katsayısı hesaplanmıştır. G çalışması sonucunda en yüksek değişkenlik kaynağının hata varyansı, ikinci en yüksek değişkenlik kaynağının öğrenci etkisi, üçüncü en yüksek değişkenlik kaynağının ise enstitü ve danışman ortak etkisinin olduğu görülmüştür. Her bir danışmanın danışmanlık yaptığı farklı öğrenci sayısı senaryoları için yapılan alternatif K çalışmaları sonucunda, öğrenci sayısının 10'un altında olduğunda güvenilirliğin düşük, 10 ile 20 arasında olduğunda güvenilirliğin orta ve 20'nin üzerinde olduğunda güvenilirliğin yüksek çıktığı gözlemlenmiştir. Klasik test kuramında çalışılan Cronbach Alfa katsayısı 0.93 oldukça yüksek çıkmıştır. Buradaki toplam varyans çoğunlukla öğrenci etkisinden kaynaklanmaktadır. Ancak araştırmanın ölçme objesinin öğrenci değil de danışman olduğu göz önüne alındığında bu değerlerin yanlış yorumlanabileceği, bu sebeple genellenebilirlik kuramıyla güvenilirlik analizinin yapılmasının daha doğru olacağı yorumu yapılmıştır.

Rae ve Hyland'ın 2001 yılında yaptıkları "Koppitz'in İnsan Figürlerini Puanlama Sisteminin Genellenebilirlik ve Klasik Test Kuramıyla Analizleri" başlıklı çalışmalarında, öğrencilerin bir insan çiziminden aldıkları puanlar genellenebilirlik ve klasik test kuramında çalışılmıştır. 8 ya da 9 yaşlarında olan 85 öğrencinin, 2 hafta arayla istedikleri tam bir insan figürünü çizmeleri ile elde edilen çizimler, 4 puanlayıcı tarafından puanlanmıştır. Genellenebilirlik kuramında tümüyle çaprazlanmış ö x p x d (öğrenci, puanlayıcı, durum) deseninde çalışılmıştır. Elde edilen sonuçlara göre; değişkenlik kaynaklarından puanlayıcı etkisi, puanlayıcı öğrenci ortak etkisi ve puanlayıcı durum ortak etkisi oldukça

düşük çıkmıştır. Bu da puanlayıcılardan kaynaklı hata varyansının oldukça düşük olduğunu göstermiştir. Klasik test kuramında test tekrar test tutarlılığı düşük, puanlayıcılar arası tutarlılık yüksek çıkmıştır. Genel olarak genellenebilirlik kuramından ve klasik test kuramından elde edilen sonuçların birbirleriyle tutarlı oldukları görülmüştür. Koppitz'in puanlama sistemine göre bir adam çiz testinden yüksek güvenilirlik elde edilmesi için öğrencilerin birkaç durumda bu testi yapmaları gerektiği vurgulanmıştır.

Sharma ve Weathers 2003 yılında yaptıkları çalışmalarında, bir ölçeğin farklı ülkelere genellenebilirliğini ve genelleme yapılabilmesi için gerekli olan madde ve birey sayısını araştırmışlardır. Araştırmaya Amerika'dan 71, Fransa'dan 70, Japonya'dan 76 ve Almanya'dan 73 birey 17 maddelik ölçeği puanlayarak katılmışlardır. Çalışmada doğrulayıcı faktör analizi ve genellenebilirlik kuramı kullanılarak sonuçlar karşılaştırılmıştır. Araştırma sonucunda, 0.90 güvenilirlik elde edebilmek için madde sayısının 11 olmasının yeterli olduğu belirtilmiştir. Yeterli birey sayısının belirlenmesi için yapılan karar çalışmaları sonucunda ise 15 maddelik bir ölçek ile 0.90'dan yüksek güvenilirlik elde edebilmek için her ülkeden 100 bireye ihtiyaç olduğu görülmüştür. Bunlara ek olarak, genellenebilirlik kuramının ve faktör analizinin ayrı ayrı avantajları ve sınırlılıklarının olduğu, her ikisinde çalışmalarda birbirini tamamlayıcı olduğu vurgulanmıştır.

Atılğan (2004), "Genellenebilirlik Kuramı ve Çok Değişkenlik Kaynaklı Rash Modelinin Karşılaştırılmasına İlişkin Bir Araştırma" başlıklı çalışmasında 2002 ve 2003 yıllarında yapılan müzik öğretmenliği özel yetenek seçme sınavı verilerini kullanarak Çok Değişken Kaynaklı Rash Modeli ve Genellenebilirlik kuramı ile kestirilen istatistikleri karşılaştırmıştır. Veriler, 2002 yılında 499 öğrenci ve 3 puanlayıcı katılımıyla 19 görev üzerinden elde edilirken, 2003 yılında 689 öğrenci ve 4 puanlayıcı katılımıyla 28 görev üzerinden elde edilmiştir. Araştırma sonucunda, Genellenebilirlik kuramı ve Çok Değişken Kaynaklı Rash Modeli ile değişkenlik kaynakları için elde sonuçların kısmen tutarlı sonuçlar ürettiği, ancak özellikle puanlayıcı ve görev değişkenlik kaynağı için tutarlı sonuçlar ürettiği görülmüştür. Bunun yanı sıra, aynı ölçme durumu için alt boyutlardan oluşan testlerde, genellenebilirlik Kuramının tek değişkenli ve çok değişkenli modellerinin farklı sonuçlar verdiği, G ve Phi katsayılarının modelden etkilendiği ve farklı puanlayıcı sayıları senaryolarının olduğu alternatif karar çalışmalarından elde

edilen G ve Phi katsayılarının, gerçek durumda kestirilen katsayılardan farklı olduğu gözlemlenmiştir.

Lee, 2005 yılında yaptığı çalışmasında TOEFL 2000 testinin konuşma bölümünün güvenilirliğini genellenebilirlik kuramıyla çalışmış ve en uygun görev ve madde sayısını alternatif K çalışmalarıyla araştırmıştır. Sınavın konuşma bölümü 6 görevden oluşan dinleme-konuşma, 2 görevden oluşan okuma-dinleme ve 5 görevden oluşan bağımsız konuşma bölümlerinden meydana gelmektedir. 2000 yılının güz döneminde TOEFL sınavının konuşma bölümünü 478 birey 13 görev üzerinden gerçekleştirmiştir. Ancak eksik doldurmalar sebebiyle araştırma 261 kişinin 11 göreve verdikleri cevaplar, 2 puanlayıcı tarafından 1 ile 5 puan arasında değerlendirilmiştir. Araştırmada genellenebilirlik kuramında (p : ö) x g, ö x g x p ve ö x g (öğrenci, puanlayıcı, görev) desenleri kullanılmıştır. Elde edilen bulgulara göre, hata varyansının çoğunluğunun puanlayıcılardan değil de teste bulunan görevlerden kaynaklandığı görülmüştür. Güvenirlik açısından görev sayısını arttırmak, puanlayıcı sayısını arttırmaktan daha etkili olmuştur. Ayrıca diğer alt bölümlere kıyasla dinleme-konuşma bölümündeki görev sayısının artırılması ile daha yüksek güvenilirlik elde edileceği görülmüştür.

Schoonen (2005), “ Yazma Puanlarının Genellenebilirliği: Bir Yapısal Eşitlik Modeli Uygulaması” başlıklı çalışmasında, kompozisyon yazma sınavlarını etkileyen değişkenlik kaynakları Genellenebilirlik kuramıyla çalışılmıştır. 442 kişilik gruptan 89 öğrenci rastlantısal olarak seçilmiş ve bu öğrencilerden 4 er kompozisyon yazmaları istenmiştir. Her kompozisyon 5 puanlayıcı tarafından, “içerik ve düzen” ve “dil kullanımını” olmak üzere 2 farklı boyutta, analitik ve bütünsel olmak üzere 2 puanlama anahtarıyla puanlanmıştır. Araştırma sonuçlarına göre, yazma sınav sonuçlarının öğrencinin yazma yeteneğinden çok değişkenlik kaynaklarından etkilendiği ve puanların genellenebilirliği açısından görev sayısını artırmanın, puanlayıcı sayısını arttırmaktan daha etkili olduğu bulunmuştur. Ayrıca, bütünsel puanlama anahtarının, analitik puanlama anahtarına göre ve dil kullanımına ilişkin puanların, içerik ve düzen puanlarına göre daha genellenebilir olduğu belirtilmiştir.

Kantor ve Lee 2007 yılında TOEFL sınavının yazma becerilerini ölçtüğü bölümün güvenilirlik analizlerini alternatif K çalışmaları yaparak genellenebilirlik kuramıyla incelemişlerdir. Alternatif K çalışmalarında en yüksek güvenilirliği elde edebilmek

için farklı görev sayısı ve farklı puanlayıcı sayısı senaryoları gerçekleştirilmiştir. Araştırma sonucunda, en yüksek güvenilirliği elde edebilmek için puanlayıcı sayısının yerine görev sayısının artırılmasının gerekli olduğu gözlemlenmiştir. Buna ek olarak, dinleme-yazma görev sayısının, okuma-yazma görev sayısından daha yüksek tutulduğunda daha yüksek güvenilirlik elde edildiği belirtilmiştir

Yelboğa (2007), “Klasik Test Kuramı ve Genellenebilirlik Kuramına Göre Güvenirliğin Bir İş Performansı Ölçeği Üzerinde İncelenmesi” başlıklı çalışmada, Genellenebilirlik kuramı ve Klasik Test Kuramı ile kestirilen istatistikleri karşılaştırmıştır. Çalışma 2005 ve 2006 yıllarında 32 maddelik bir iş performansı ölçeği ile 176 çalışanın 3 puanlayıcı tarafından değerlendirilmesi sonucu elde edilen veriler kullanılarak gerçekleştirilmiştir. Klasik Test Kuramı kapsamında test-tekrar-test, Cronbach Alfa güvenilirlik katsayıları ve puanlayıcılar arası güvenilirlik için Kendall’ın uyum katsayısı hesaplanmıştır. Genellenebilirlik kuramı kapsamında ise çok değişkenlik kaynaklı modellerde G ve Phi katsayıları hesaplanmıştır. Araştırma sonuçlarına göre, Genellenebilirlik kuramı ve Klasik test kuramına göre elde edilen güvenilirlik katsayılarının birbirleriyle tutarlı olduğu görülmüştür. Ayrıca iş performansı ölçeğinin güvenilirliği yüksek bir ölçme aracı olarak kabul edilebileceği ve benzer çalışmalarda gerek alternatif karar çalışmaları gerekse tek bir analizle güvenilirlik katsayısına ulaşılabilmesi nedeniyle G kuramının kullanılması önerilmiştir.

Güler’in (2008) yaptığı “Klasik Test Kuramı, Genellenebilirlik Kuramı ve Rasch Modeli Üzerine Bir Araştırma” başlıklı çalışmada, TIMMS 1999 açık uçlu matematik sorularından 24’ü kullanılarak elde edilen veriler ile Klasik Test Kuramı, Genellenebilirlik kuramı ve Çok Değişken Kaynaklı Rasch Modeli ile incelenmiştir. Veriler 2007 bahar öğretim yılında dört puanlayıcının 203 öğrencinin cevaplarını bütünsel dereceli puanlama anahtarı ile puanlaması sonucu elde edilmiştir. Klasik test kuramında Cronbach alfa iç tutarlılık katsayısı, puanlayıcılar arası korelasyon katsayısı ve puanlayıcılar arası güvenilirlik için Kendall’ın uyum katsayısı, genellenebilirlik kuramındaki tümüyle çaprazlanmış b x g x p desende G ve Phi katsayıları hesaplanmıştır. Rasch modelinde ise birey, görev ve puanlayıcı boyutlarına ilişkin ayrı ayrı güvenilirlik hesaplamaları yapılmıştır. Araştırma sonucunda, puanlayıcıların verdikleri puanların ortalamaları arasında farklılık olmasına rağmen, birbirleriyle tutarlı puanlamalar yaptıkları sonucuna varılmıştır.

Ayrıca, performans ölçülmesinde bu üç kuramdan hangisinin kullanılacağına belirlenmesinde elde edilen puanların hangi ölçek yapısına uygun olduğu, ölçmenin hangi amaçla yapılacağı ve verilerin kullanılacak kuramın sayıltılarını sağlayıp sağlamadığının göz önünde bulundurulması gerektiği vurgulanmıştır. Bu tarz çalışmalarda az iki kuramın birlikte kullanılması önerilmiştir.

Atılğan (2008), 2003-2004 öğretim yılında İnönü Üniversitesinde 249 adayın 4 puanlayıcı tarafından puanlandığı müzik öğrencilerinin özel yetenek seçme sınav sonuçlarının güvenilirlik analizini yapmıştır. Güvenirlik analizleri, sınav 3 alt bölümden (çözümleme, şarkı söyleme ve çalma) oluştuğu ve birden fazla puanlayıcı tarafından puanlandığı için genellenebilirlik kuramında tümüyle çaprazlanmış  $b \times g \times p$  deseniyle yapılmıştır. Çalışmanın amacı, özel yetenek seçme sınavının güvenilirliğini hesaplamak, değerlendirme için en uygun görev ve puanlayıcı sayısını belirlemektir. Bu amaçla puanlayıcı sayısının 1 artırılıp azaltılması ve görev sayısının 2 artırılıp azaltılması ile alternatif karar çalışmaları yapılmıştır. Araştırma sonuçlarına göre, hem her bir bölüm için hem de testin tümü için, puanlayıcı ve görev sayısının artırılmasının G ve Phi katsayılarında önemli ölçüde artışlara neden olmadığı belirtilmiştir. Bu sebeple, testin daha fazla görev ile 3 puanlayıcı tarafından puanlanmasının daha ekonomik olacağı sonucuna varılmıştır.

Au, Prahardi ve Shiell (2008) yaptıkları çalışmada 1990 ve 2003 yılları arasında yayınlanan 30 makaleyi, 2 puanlayıcıya birbirlerinden bağımsız olarak 2 ölçek kapsamında puanlatmış, bu iki ölçeğin güvenilirliğini ve hangisinin daha kullanışlı olduğunu araştırmışlardır. Puanlama için kullanılan ölçeklerden birisi 16 maddelik Likert tipi bir ölçek, diğeri ise 57 maddelik bir kontrol listesidir. Elde edilen veriler klasik test kuramında sınıf içi korelasyon katsayısı ile, genellenebilirlik kuramında ise tümüyle çaprazlanmış olan  $m \times p \times ö$  (makale, puanlayıcı, ölçek) deseniyle güvenilirlik analizleri yapılmıştır. Çıkan bulgulara göre, her iki kuram için de puanlayıcılar arası güvenilirlik ve ölçeklerin güvenilirlikleri oldukça yüksek çıkmıştır. Bunun sonucunda ise, ilk ölçeğin daha az maddeden oluştuğu ve daha az zaman harcadığı için, ikinci ölçeğin ise daha detaylı bilgi verdiği için kullanımının avantajlı olduğu belirtilmiştir.

Nalbantoğlu (2009), "Performans Ölçümlerinde Genellenebilirlik Kuramıyla Farklı Desenlerin Karşılaştırılması" başlıklı çalışmasında OSCE (objektif yapılandırılmış

linik sınav) sınavının güvenilirliğini Genellenebilirlik kuramında puanlayıcıların birlikte ve dönüşümlü olarak puanlamasıyla oluşturulan  $\bar{o} \times g \times p$  ve  $(\bar{o} : g) \times p$  desenleri için ayrı ayrı G ve K çalışmaları yapmış ve sonuçları karşılaştırmıştır. Araştırmada 2007-2008 öğretim yılında Hacettepe Üniversitesi Tıp Fakültesi 3. Sınıf öğrencilerinden tesadüfi olarak seçilen 48 öğrencinin iletişim becerileri istasyonundaki hasta görüşmeleri 3 puanlayıcı tarafından puanlanmıştır. Araştırma sonucunda,  $\bar{o} \times g \times p$  ve  $(\bar{o} : p) \times g$  desenlerinde değişkenler için kestirilen varyans değerlerinin birbirleriyle paralellik gösterdiği, puanlayıcıların öğrencileri puanlamalarında farklılıklar olmadığı görülmüştür. Ayrıca, her iki desen için de yapılan karar çalışmaları sonucunda elde edilen G ve Phi katsayıları arasında çok fark olmadığı, ancak  $(\bar{o} : p) \times g$  deseninde katsayıların daha büyük çıkma eğiliminde olduğu belirtilmiştir.

Deliceoğlu'nun (2009) "Futbol Yetilerine İlişkin Dereceleme Ölçeğinin Genellenebilirlik ve Klasik Test Kuramına Dayalı Güvenirliklerinin Karşılaştırılması" adlı çalışmasında 56 maddeden oluşan ölçek 4 puanlayıcı tarafından 72 futbolcu için puanlanmıştır. Elde edilen veriler klasik test kuramında test-tekrar-test tutarlılığı için Pearson Momentler Çarpım korelasyon katsayısı, maddelerin iç güvenilirliği için Cronbach Alfa katsayısı ve puanlayıcılar arası güvenilirlik için Kendall'ın uyuşum katsayısı hesaplanmıştır. Genellenebilirlik kuramında ise çok değişkenli  $b \times m \times p$  modeli için G ve Phi katsayıları bulunmuştur. Elde edilen bulgulara göre, G katsayısı ve Cronbach Alfa katsayısının beklenen değerlerinden yüksek olduğu, Phi katsayısı ve Kendall'ın uyuşum katsayısının ise beklenenden düşük çıktığı gözlemlenmiştir. Ayrıca, Futbol Yetilerine İlişkin Dereceleme Ölçeğinin her iki kuramda da güvenilirlik katsayılarının yüksek olması nedeniyle güvenilir bir ölçme aracı olarak kabul edilebileceği belirtilmiştir.

Christ ve diğerleri 2010 yılında yaptıkları çalışmalarında 125 puanlayıcı, 6 öğrencinin çözülemeyen bir lego ile uğraşmalarını içeren 3'er dakikalık videoları puanlamışlardır. Puanlama işlemi 3 ayrı durumda 6'lı, 10'lu ve 14'lü noktalanmış 60 milimetrelilik metrik ölçekle gerçekleştirilmiştir. Araştırmada genellenebilirlik kuramında tümüyle çaprazlanmış  $\bar{o} \times p \times d$  (öğrenci, puanlayıcı, durum) deseni ile çalışılmıştır. Araştırma sonucunda gözlenen durum sayısının artırılmasının ve 10'lu noktalandırılmış metrik ölçeğin kullanımının daha geçerli sonuçlar verdiği belirtilmiştir.

Öztürk (2011), “Voleybol Becerileri Gözlem Formu İle Elde Edilen Puanların Genellenebilirlik ve Klasik Test Kuramına Göre Karşılaştırılması” başlıklı çalışmada 25 maddeden oluşan gözlem formu, 102 erkek voleybolcu için 4 puanlayıcı tarafından doldurulmuştur. Elde edilen veriler, klasik test kuramında puanlayıcı iç tutarlılığı için Cronbach Alfa katsayısı ve puanlayıcılar arası güvenilirlik için Kendall’in uyuşum katsayısı hesaplanmıştır. Genellenebilirlik kuramında ise tümüyle çaprazlanmış  $b \times g \times p$  modeli için genellenebilirlik ve güvenilirlik katsayıları bulunmuştur. Farklı puanlayıcı ve görev sayısı senaryolarıyla oluşturulan karar çalışmalarında ana ve ortak etkilerin varyans bileşenlerinin kestirilmesi için G ve Phi katsayıları hesaplanmış ve en büyük varyans bileşeninin  $g \times p$  ortak etkisi varyans bileşeni olduğu bulunmuştur. Klasik test kuramından ve genellenebilirlik kuramından elde edilen güvenilirlik katsayılarının birbirleriyle tutarlı oldukları görülmüştür. Ancak, güvenilirlik katsayıları incelendiğinde G ve Phi katsayıları, Cronbach Alfa katsayısı ve Kendall’in uyuşum katsayısı beklenen değerlerinden düşük olduğu belirtilmiştir. Bu sebeple, Voleybol Becerilerine İlişkin Dereceleme Ölçeğinin güvenilir bir ölçme aracı olarak değerlendirilemeyeceği vurgulanmıştır.

Büyükkıdık (2012), “Problem Çözme Becerisinin Değerlendirilmesinde Puanlayıcılar Arası Güvenirliğin Klasik Test Kuramı Ve Genellenebilirlik Kuramına Göre Karşılaştırılması” başlıklı çalışmasını, 132 öğrencinin matematik problem çözme becerisini ölçmeye yönelik hazırlanan 2 adet performans görevinde gösterdikleri performansların, dört puanlayıcı tarafından analitik ve bütünsel dereceli puanlama anahtarıyla puanlanması ile gerçekleştirmiştir. Klasik test kuramında puanlayıcılar arası güvenilirlik sınıf içi ilişki katsayısı ve puanlayıcılar arası ilişki katsayısı ile hesaplanmıştır. Aynı puanlayıcıların, aynı performanslara analitik ve bütünsel puanlama anahtarı kullanarak verdikleri puanların ortalamaları arasında fark olup olmadığı bağımlı gruplar t testi ile belirlenmiştir. Genellenebilirlik kuramında ise  $b \times g \times p$  ve anahtar çeşidinin değişkenlik kaynağı olarak ele alındığı  $b \times g \times p \times a$  tümüyle çaprazlanmış desenlerde G ve Phi katsayıları hesaplanmıştır. Elde edilen bulgulara göre, klasik test kuramından elde edilen güvenilirlik katsayılarının, genellenebilirlik kuramından elde edilen güvenilirlik katsayılarına göre göreceli olarak daha düşük olduğu görülmüştür. Genellenebilirlik kuramının, klasik test kuramına göre daha detaylı bilgi verdiği vurgulanmıştır. Ayrıca, dereceli puanlama anahtarı türünün de ölçümlerin



güvenirliğine etki edebileceği belirtilmiştir. Bu araştırmada her iki puanlama anahtarı ile hesaplanan güvenilirlik katsayıları yüksek olsa da, analitik dereceli puanlama anahtarından elde edilen puanların daha yüksek güvenilirliğe sahip olduğu görülmüştür.

Alkan (2013), “ PISA 2009 Okuma Becerileri Açık Uçlu Sorularının Puanlanmasında Genellenebilirlik Kuramındaki Farklı Desenlerin Karşılaştırılması” başlıklı çalışmasında 886 öğrencinin cevaplarının, birden fazla puanlayıcı tarafından birlikte ve dönüşümlü olarak puanlanmasıyla elde edilen  $\bar{o} \times s \times p$  ve  $(\bar{o} : p) \times s$  desenleri genellenebilirlik kuramına göre karşılaştırmıştır. Araştırma sonuçlarına göre, öğrenci ve puanlayıcı değişkenlerinin yuvalanmış olduğu, soru değişkenlerinin ise çaprazlanmış olduğu  $(\bar{o} : p) \times s$  deseninden elde edilen G ve Phi katsayıları, tümüyle çaprazlanmış olan  $\bar{o} \times s \times p$  deseninden elde edilen G ve Phi katsayılarına göre daha yüksek olduğu görülmüştür. Ayrıca, her iki desende yapılan alternatif karar çalışmaları incelendiğinde, puanlayıcı sayısının artırılmasının G ve Phi katsayılarında artışa sebep olduğu gözlemlenmiştir.

Eser ve Gelbal (2013), 106 öğrencinin 3 puanlayıcı tarafından dereceli puanlama anahtarı kullanılarak gerçekleştirilen performans puanlamasında ortaya çıkan puanlayıcı tutarlılığını, genellenebilirlik kuramı ve lojistik regresyon analizinden yararlanarak karşılaştırmışlardır. Analizler, hem madde bazında hem de testin tamamına dayalı olarak gerçekleştirilmiştir. Madde bazında yapılan incelemelerde genellenebilirlik kuramında  $\bar{o} \times p$  deseninde analiz edilip, puanlayıcı varyans bileşenleri ve toplam varyansı açılama yüzdeleri hesaplanmıştır. Lojistik regresyon analizi ile de öğrencilerin her bir maddeden aldıkları puanlara göre sınıflama yüzdeleri belirlenmiştir. Testin tamamına dayalı yapılan analizlerde ise genellenebilirlik kuramında  $\bar{o} \times m \times p$  deseni uygulanıp, lojistik regresyon analizi ile de öğrencilerin her bir puanlayıcıdan aldıkları toplam puana göre başarılı-başarısız sınıflamaları yapılmıştır. Elde edilen bulgulara dayalı olarak puanlayıcılar arası tutarlılığı belirlemede genellenebilirlik kuramı ile lojistik regresyon analizinin paralel sonuçlar verdiği gözlemlenmiştir. Ancak, genellenebilirlik kuramının lojistik regresyon analizine göre daha detaylı bilgiler vermesi nedeniyle, lojistik regresyon analizinin genellenebilirlik kuramına göre daha yüzeysel bir istatistik olduğu belirtilmiştir.

Yıldıztekin'in (2014) "Klasik Test Kuramı ve Genellenebilirlik Kuramından Puanlayıcılar Arası Tutarlılığın Farklı Yöntemlere Göre Karşılaştırılması" başlıklı çalışmasında, 84 öğrencinin matematik problem çözme becerisini ölçmeye yönelik hazırlanan 6 adet açık uçlu sorudan oluşan bir teste verdikleri cevaplar, beş puanlayıcı tarafından analitik ve bütünsel dereceli 2 ayrı puanlama anahtarı kullanılarak 20-25 gün arayla puanlanmıştır. Elde edilen verilerin, Klasik test kuramı kapsamında Pearson momentler çarpım korelasyon katsayısı, Spearman sıra farkları korelasyon katsayısı, Cronbach Alfa, Kappa ve Krippendorff Alfa katsayıları ile güvenilirlik analizleri yapılmıştır. Genellenebilirlik kuramında ise b x m x p modeli çalışılmıştır. Araştırma sonuçlarına göre, KTK ve GK'dan elde edilen güvenilirlik katsayılarının birbirleriyle tutarlı ve oldukça yüksek oldukları görülmüştür. Genellenebilirlik kuramının b x m x p deseninden elde edilen sonuçlara göre, puanlayıcıların analitik ve bütünsel puanlama anahtarlarıyla verdikleri puanlar arasında değişkenliğe etki etmedikleri belirlenmiştir. Ayrıca, puanlayıcılar arası tutarlılığın yüksek olduğu belirlenmiştir.

Chen ve arkadaşları 2014 yılında yaptıkları araştırmalarında, okul öncesi eğitim programı değerlendirme ölçeğinin (Chinese Early Childhood Program Rating Scale, CECPRS) güvenilirliğini genellenebilirlik kuramıyla araştırmışlardır. Okul öncesi eğitim programı değerlendirme ölçeği, mekân ve mobilya, kişisel bakım rutini, müfredat planlama ve uygulama, grup çalışması, aktiviteler, dilsel akıl yürütme, rehberlik ve iletişim, aile ve çalışanlar olmak üzere 8 alt boyuttan oluşan 51 maddelik 7'li Likert tipi bir ölçektir. Araştırmada Çin'de bulunan 91 anaokulundan, 176 anaokulu sınıfı rastlantısal olarak seçilmiş ve bu sınıflar okul öncesi eğitim programı değerlendirme ölçeği ile 2 puanlayıcı tarafından puanlanmıştır. Puanlayıcılar, 4 günlük alan çalışması sonucunda 0,85 lik puanlayıcılar arası tutarlılığı yakaladıktan sonra gerçek sınıfları puanlamışlardır. Araştırma sonuçlarında, en yüksek güvenilirlik (0,96) mekân ve mobilya alt ölçeği için elde edilirken, en düşük güvenilirlik (0,91) rehberlik ve iletişim alt ölçeği için elde edilmiştir. Farklı puanlayıcı sayısı senaryoları ile gerçekleştirilen karar çalışmalarında, puanlayıcı sayısını artırmanın güvenilirlik katsayısında ciddi bir artışa sebep olmadığı, 2 puanlayıcı ile yeterli düzeyde güvenilirlik sağlandığı belirtilmiştir.

## 2.1. İlgili Arařtırmalar Özet

Yurt içinde yapılan alıřmalara iliřkin olarak genellikle yalnız klasik test kuramı ya da yalnız genellenebilirlik kuramının alıřıldığı görülmüřtür. Her iki kuramdan elde edilen bilgilerin karşılařtırıldığı daha az sayıda alıřma bulunmaktadır. Klasik test kuramı ve genellenebilirlik kuramının karşılařtırıldığı alıřmalarda, Klasik test kuramına göre farklı teknikler uygulanırken (Cronbach alfa katsayısı, Pearson moment arpım korelasyon katsayısı, Kendall'in uyuşum katsayısı, sınıf içi iliřki katsayısı, Kappa, Krippendorf Alfa vs.), Genellenebilirlik kuramında genellikle tümüyle aprazlanmış desen alıřılmıştır. Ayrıca, Klasik test kuramı, Genellenebilirlik kuramı ve Rasch modelinin karşılařtırıldığı alıřmalarda mevcuttur. Bu alıřmalardan farklı olarak, Genellenebilirlik kuramında farklı desenlerin karşılařtırıldığı alıřmalar da bulunmaktadır. Bu alıřmalarda tümüyle aprazlanmış desen ile yuvalanmış desenin karşılařtırılıp, hangi desenden daha çok bilgi sađlandığı, daha yüksek güvenilirlik elde edildiđi ve özellikle birden fazla puanlayıcının olduđu alıřmalarda hangi desen ile en ekonomik alıřmanın yapıldığı arařtırılmıştır.

Yurt dıřında yapılan arařtırmalar incelendiđinde, daha çok Genellenebilirlik kuramına göre farklı senaryolara iliřkin karar alıřmalarının yapıldığı, en uygun madde sayısı, puanlayıcı sayısı ve uygun örneklem büyüklüklerinin belirlenmeye alıřıldığı görülmüřtür.

Genel olarak alanyazın incelendiđinde, Klasik test kuramına göre uygulanan tekniklerin basit ve kolay yorumlanabilir olması avantaj kabul edilirken, birden fazla deđişkenlik kaynađının olduđu durumlarda tek bir hata kaynađına göre güvenilirlik hesaplaması en büyük sınırlılıđı olarak deđerlendirilmektedir. Genellenebilirlik kuramının ise birden fazla deđişkenlik kaynađını birlikte göz önünde bulundurarak güvenilirliđi belirlemesi ve karar alıřmaları ile farklı sayıdaki deđişkenlik kaynakları için önceden bilgi vermesi açısından Klasik test kuramına göre daha çok tercih edildiđi görülmektedir.

### 3. YÖNTEM

Bu bölümde araştırmanın türü, araştırma grubu, ver toplama aracı ve verilerin analizi yer almaktadır.

#### 3.1. Araştırmanın Türü

Araştırma, puanlayıcılar arası güvenilirliğin hesaplanmasında Klasik Test Kuramı ve Genellenebilirlik Kuramının avantaj ve dezavantajlarını ortaya koyması, kuramlardan hangisinin daha çok bilgi sağladığını belirlemesi ve iki kuramdan elde edilen sonuçları karşılaştırması açısından bu yönüyle karşılaştırmalı bir araştırmadır. Aynı zamanda araştırma, Genellenebilirlik kuramı ve Klasik Test Kuramı ile Otizm Sosyal Beceriler Profili ölçeğine ait özelliklerin belirlenmesi yönüyle durum belirleme çalışması olduğundan betimsel bir araştırma niteliği taşımaktadır.

#### 3.2. Araştırma Grubu

Araştırmanın çalışma grubunu, Ankara'da bir özel eğitim ve rehabilitasyon merkezinde eğitim almakta olan, "Otizm", "Asperger Sendromu", "Başka Şekilde Tanımlanamayan" tanılarını almış ya da bu gruplardan herhangi birine dahil edilemeyen ancak otizm spektrum bozukluğu belirtileri gösteren "Yaygın Gelişimsel Bozukluk" tanısı almış 6-17 yaş arasında olan 50 çocuk ve genç oluşturmuştur. Örneklem grubu seçilirken, puanlayıcıların en az 1 yıldır birlikte çalıştıkları öğrenciler seçilmiştir.

**Tablo 3.1: Öğrenci Betimsel Bilgileri ve Frekansları**

<i>Cinsiyet</i>	<i>Frekans</i>	<i>Yüzde (%)</i>
<i>Kız</i>	13	26
<i>Erkek</i>	37	74
<i>Toplam</i>	50	100

50 öğrenci, Otizm Sosyal Beceriler Profili ölçeğinin "sosyal karşılıklılık" alt boyutunda bulunan 15 madde ile 5 puanlayıcı tarafından puanlanmıştır. Puanlayıcı grubunu ise aynı kurumda 3-5 yıldır görev yapmakta olan 2 özel eğitim öğretmeni, 1 psikolog (özel eğitim alanıyla ilgilenen), 1 fizyoterapist ve 1 sosyal psikolog (aynı zamanda davranış terapisti) oluşturmaktadır.

### 3.3. Veri Toplama Aracı

Bu çalışmada, Türkçe uyarlaması yapılan Otizm Sosyal Beceriler Profili ölçeğinin “sosyal karşılıklılık” alt ölçeğinde yer alan 15 madde kullanılmıştır. Alt ölçek Ek-4’te yer almaktadır. Ayrıca, çalışma öncesinde çalışmaya katılan öğrenciler için veli onay formu ve puanlayıcılar için gönüllü katılım formları hazırlanmıştır. Bu formlar çalışmanın gönüllülük esasına dayandığını, katılımcı kimliklerinin gizli kalacağını ve çalışmanın istenildiği zaman bırakılabileceğini garanti etmektedir. Bu formlar da EK-5 ve EK-6’da yer almaktadır.

#### 3.3.1. Otizm Sosyal Beceriler Profili (OSBP)

OSBP, 2007 yılında Bellini ve Hopf tarafından otizmlili çocukların sosyal beceri yetersizliklerinin belirlenmesi ve bu yetersizliklere uygun müdahale programlarının oluşturulması amacıyla geliştirilmiştir. Araştırmanın örneklemini, otizm spektrum bozukluğuna sahip 6-17 yaş arası 340 çocuk ve genç oluşturmuştur. 45 maddelik Otizm Sosyal Beceriler Profili 3 alt boyuttan meydana gelmektedir. Bu alt boyutlar; sosyal karşılıklılık, sosyal katılım/kaçınma ve zarar verici sosyal davranışlar şeklinde adlandırılmıştır. 4’lü Likert tipi dereceleme ölçeği olan OSBP de her madde için “hiçbir zaman” seçeneği 1 puan, “bazen” seçeneği 2 puan, “sık sık” seçeneği 3 puan, “her zaman” seçeneği 4 puan ile değerlendirilmektedir.

Ölçeğin güvenirliği Cronbach Alfa katsayısıyla hesaplanmış ve ölçeğin iç tutarlığı  $\alpha=0.92$  bulunmuştur. Alt ölçekler için güvenirlik katsayıları ise sosyal karşılıklılık için  $\alpha=0.92$ , sosyal katılım/kaçınma için  $\alpha=0.89$  ve zarar verici sosyal davranışlar için  $\alpha=0.84$ ’tür. Alt ölçeklerin birbirleriyle ilişkileri  $r=0.09$  ile  $r=0.36$  arasında bulunmuş ve bu durum yazarlar tarafından alt ölçeklerin birbirlerinden bağımsız olmasıyla açıklanmıştır. Ölçeğin 1 hafta arayla 2 kez puanlanması sonucu elde edilen verilere göre alt ölçeklerin test-tekrar-test güvenirliği sosyal karşılıklılık için 0.89, sosyal katılım/kaçınma için 0.86, zarar verici sosyal davranışlar için 0.84 ve ölçeğin tamamı için 0.90 olarak bulunmuştur.

#### 3.3.2. OSBP Türkçe Formu (OSBP-T)

OSBP’nin Türkçe uyarlamasının ilk aşamasını, çeviri çalışmaları oluşturmuştur. İlk olarak ölçek, Türkçe ve İngilizceyi iyi düzeyde bilen üç uzman tarafından birbirlerinden bağımsız olarak Türkçeye çevrilmiştir. Çevirilerde uyumsuzluk yaşanan maddeler belirlendikten sonra, ölçek tekrar üç uzman tarafından

İngilizceye çevrilmiştir. Çevirinin son aşamasında iki dile de iyi derecede hâkim, özel eğitim alanında uzman iki bağımsız kişi tarafından ölçeğin Türkçe ve İngilizce çevirileri karşılaştırılarak gerekli düzeltmeler yapılmış ve ölçeğe son hali verilmiştir.

Otizm Sosyal Beceriler Profili Türkçe formunun geçerlik ve güvenilirlik çalışmaları 208 otizmlı çocuk ve gencin ebeveynleri tarafından değerlendirilmesi sonucu elde edilen verilerle gerçekleştirilmiştir. İlk olarak ölçeğin yapı geçerliğini belirlemek için doğrulayıcı faktör analizi (DFA) yapılmıştır. Bellini ve Hopf (2007) çalışmalarında ölçeğin faktör yapısını belirledikleri için önce doğrulayıcı faktör analizi yapılmıştır. Ancak hesaplanan uyum indeksleri kabul edilebilir düzeyde olmadığı için açıklayıcı faktör analizi (AFA) yapılmasına karar verilmiştir. Yapılan analiz sonucunda, faktör yük değerleri 0.30'dan yüksek olup, iki veya daha fazla faktörde 0.10'dan düşük farka sahip maddeler ile faktör yük değeri 0.30'un altında olan 4 maddenin ölçekten çıkarılmasına karar verilmiştir. Böylece ölçeğin Türkçe formunda 45 madde yerine 41 madde yer almıştır. Ölçeğin 3 faktörlü bir yapıya sahip olduğu, Varimax dik döndürme tekniği kullanılarak belirlenmiş, bu üç faktörün toplam varyansın 0.46'sını açıkladığı görülmüştür. Açıklayıcı faktör analizi ile elde edilen 3 faktörlü yapının sınanması için doğrulayıcı faktör analizi uygulanmıştır. Yapılan analizler sonucunda, bazı uyum indekslerinin düşük olmasına karşın ölçeğin üç faktörlü yapısının genel olarak kullanılabilir olduğu söylenmiştir.

Daha sonra ölçeğin ölçüt geçerliğinin belirlenmesi için 2 ayrı çalışma yapılmıştır. İlk çalışmada, ölçekten elde edilen puan ortalamaları bilinen grupların ortalamasıyla karşılaştırılmıştır. Bunun için zihinsel engelli tanısı konmuş öğrencilerin öğretmenlerinden OSBP-T'nin bu çocuk ve gençler için doldurulması istenmiş, elde edilen sonuçlar otizmlı çocuk ve gençlerin aldıkları puanlarla karşılaştırılmış ve aralarında beklendiği üzere anlamlı bir fark bulunmuştur. Ölçüt geçerliğinin ikinci çalışmasında ölçeğin benzer yapıları ölçen Sosyal Beceri Dereceleme Sistemi (SBDS) ile toplanan verilerden elde edilen puanlarla ilişkisine (Özgüven, 2000) bakılmış ve orta derecede bir ilişki ( $r = .69, p < .01$ ) bulunmuştur. Yapılan yapı ve ölçüt geçerliği çalışmaları sonucunda OSBP'nin ülkemizde kullanılabilir geçerli bir araç olduğuna karar verilmiştir.

Son olarak ölçeğin güvenilirlik çalışmaları yapılmıştır. Bu aşamada ölçeğin iç tutarlılığını belirlemek amacıyla Cronbach Alfa katsayıları hesaplanmıştır. Alt

ölçekler için güvenilirlik katsayıları; sosyal karşılıklık için  $\alpha=0.91$ , sosyal katılım/kaçınma için  $\alpha=0.91$ , zarar verici sosyal davranışlar için  $\alpha=0.78$  ve ölçeğin tamamı için  $\alpha=0.84$  bulunmuştur. Bu sonuçlara göre Otizm Sosyal Beceriler Profili ölçeğinin güvenilir bir araç olduğuna karar verilmiştir.

### 3.4. Verilerin Analizi

4'lü Likert tipi bir ölçek olan Otizm Sosyal Beceriler Profili kullanılarak 50 öğrenci, 5 uzman tarafından değerlendirilmiştir. Puanlayıcılar tamamen birbirlerinden bağımsız puanlama yapmışlardır. Elde edilen veriler, klasik test kuramı ve Genellenebilirlik kuramındaki tamamen çaprazlanmış  $b \times m \times p$  (birey, madde, puanlayıcı) deseni ve birey ile puanlayıcıların yuvalandığı, maddelerin ise çaprazlandığı yuvalanmış desen olan  $(p : b) \times m$  deseniyle analiz edilmiştir. Verilerin çözümlenmesi 6 aşamada gerçekleştirilmiştir. Buna göre;

➤ Birinci aşamada, öğrenci puanlarının her bir puanlayıcıya ilişkin betimsel istatistikleri hesaplanmıştır.

Klasik test kuramı kapsamında;

➤ İkinci aşamada, Otizm Sosyal Beceriler Profili ölçeğindeki maddelerin kendi içinde tutarlı ölçme yapıp yapmadığının belirlenmesi amacıyla her bir puanlayıcı için iç tutarlılık güvenilirliğini ifade eden Cronbach Alfa ( $\alpha$ ) katsayıları hesaplanmıştır. Daha sonra, beş farklı puanlayıcının puanlamaları arasındaki tutarlılık düzeyinin belirlenmesi için parametrik olmayan istatistiksel bir teknik olan Kendall'in uyuşum katsayısı (Kendall's Concordance Coefficient) hesaplanmıştır. Ayrıca, her bir puanlayıcının verdiği puanlar ile diğer bir puanlayıcının verdiği puanlar arasındaki ilişki, Pearson Momentler Çarpım korelasyon katsayısı ile analiz edilmiştir. Korelasyon katsayılarının ortalamadan bağımsız olarak hesaplanması ve puanlayıcıların verdikleri puanlar arasındaki benzerlik ve farklılıklar hakkında bilgi vermemesi nedeniyle, elde edilen puanların ortalamaları arasında anlamlı bir fark olup olmadığı ilişkili örneklem için tek değişkenli varyans analizi ile test edilmiştir.

Genellenebilirlik kuramı kapsamında;

➤ Üçüncü aşamada, birey (b), madde (m) ve puanlayıcı (p) değişkenlerinin çapraz tasarlandığı  $b \times m \times p$  deseni çalışılmıştır. Bu aşamada ilk olarak, G çalışması sonucunda kestirilen varyans bileşenleri ve toplam varyansı açıklama

yüzdeleri incelenmiştir. Daha sonra, puanlayıcı sayılarının 3, 4, 6 ve 7 olduğu, madde sayılarının ise 10, 20, 25 ve 30 olduğu senaryolar için yapılan karar çalışmaları sonucu elde edilen G ve Phi katsayıları bulunmuştur.

➤ Dördüncü aşamada, birinci desende kullanılan aynı verilerle birey (b) ve puanlayıcı (p) değişkenlerinin yuvalanmış, madde (m) değişkeninin ise çapraz tasarlandığı (p : b) x m deseni çalışılmıştır. Bu aşamada ilk olarak, G çalışması sonucunda kestirilen varyans bileşenleri ve toplam varyansı açıklama yüzdeleri incelenmiştir. Daha sonra karar çalışmalarında, puanlayıcı sayılarının 3, 4, 6 ve 7 olduğu, madde sayılarının ise 10, 20, 25 ve 30 olduğu senaryolara ilişkin G ve Phi katsayılarının değişimine bakılmıştır.

➤ Beşinci aşamada, birey (b), madde (m) ve puanlayıcı (p) değişkenlerinin çapraz tasarlandığı b x m x p deseni ve birey (b) ve puanlayıcı (p) değişkenlerinin yuvalanmış, madde (m) değişkeninin ise çapraz tasarlandığı (p : b) x m deseninden elde edilen G çalışması sonuçlarının karşılaştırılması yapılmıştır.

➤ Altıncı aşamada, b x m x p ve (p : b) x m desenlerinde puanlayıcı ve madde sayılarının artırılıp azaltılmasıyla yapılan karar çalışmaları ile her iki desenden elde edilen G ve Phi katsayıları karşılaştırılmıştır.

Yapılan analizlerde SPSS 20.0 ve EduG 6.1 programlarından yararlanılmıştır.



## 4. BULGULAR VE TARTIŞMA

Bu bölümde, Otizm Sosyal Beceriler Profili ölçeğinin “sosyal karşılıklık” alt boyutundan elde edilen puanlamalara göre, araştırmanın her bir alt problemine ilişkin bulgu ve bulgulara dayalı olarak yapılan yorumlara yer verilmektedir. Bulgu ve yorumlardan önce, uygulamadan elde edilen puanların betimsel istatistikleri yer almaktadır.

### 4.1. Betimsel İstatistikler

15 madde doğrultusunda 5 puanlayıcının ayrı ayrı 50 bireye verdikleri puanların betimsel istatistikleri Tablo 4.1’de verilmiştir.

**Tablo 4.1. Elde Edilen Puanların 5 Puanlayıcıya İlişkin Betimsel İstatistikleri**

<i>İstatistikler</i>	<i>Puanlayıcılar</i>				
	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>
<i>Minimum</i>	15	15	15	17	15
<i>Maksimum</i>	59	56	60	53	55
<i>Ortalama</i>	33.98	30.68	33.38	29.88	23.94
<i>Std. Sapma</i>	12,28	13,00	12,73	10,06	9,41
<i>Çarpıklık</i>	-,023	,312	-,181	,654	1,332
<i>Basıklık</i>	1,332	-1,258	-1,263	-,437	1,446

Tablo 4.1 incelendiğinde, en yüksek ortalamanın 1. puanlayıcıya (33.98), en düşük ortalamanın ise 5. puanlayıcıya (23.94) ait olduğu görülmektedir. Tüm puanlayıcıların yaptıkları puanlamalar ile elde edilen çarpıklık ve basıklık katsayıları -2 ile +2 arasında değişmektedir. Birinci ve üçüncü puanlayıcı için çarpıklık katsayısı negatif değerler almıştır. Negatif değerler dağılımın sola çarpık olduğunu ifade etmektedir. İkinci, dördüncü ve beşinci puanlayıcılar için ise çarpıklık katsayısı pozitif değerler almıştır. Pozitif değerler ise dağılımın sağa çarpık olduğunu ifade etmektedir. Basıklık katsayılarına bakıldığında, ikinci, üçüncü ve dördüncü puanlayıcıların verdiği puanların basıklık katsayıları negatif değerler alıp, puanlar normalden daha basık; diğer puanlayıcıların verdikleri puanların basıklık katsayıları ise pozitif değerler alıp, puanlar normalden daha sivri dağılım göstermektedir.

#### 4.2. Alt Problem 1: Otizm Sosyal Beceriler Profilinden Elde Edilen Puanların Klasik Test Kuramına Göre İncelenmesi

Bu bölümde, “Otizm Sosyal Beceriler Profilinden elde edilen puanların 5 puanlayıcıya ilişkin iç tutarlılık düzeyi nedir?” ve “Otizm Sosyal Beceriler Profiline ilişkin değerlendirmede 5 farklı puanlayıcı arasındaki tutarlılık derecesi nedir?” sorularına yanıt aranmıştır.

##### 4.2.1. Klasik Test Kuramına Göre Otizm Sosyal Beceriler Profilinden Elde Edilen Puanların 5 Puanlayıcıya İlişkin İç Tutarlılık Düzeyleri

Ölçekteki maddelerin birbirleriyle korelasyonu anlamına gelen iç tutarlılık, ölçekteki maddelerin tek bir yapıyı ölçme derecesidir (Cohen ve Swerdlick, 2009). Aynı zamanda bu iç tutarlılık katsayısı, bir puanlayıcının puanlarının kendi içinde tutarlı olup olmadığı hakkında da bilgi vermektedir. Bu sebeple, Otizm Sosyal Beceriler Profili ölçeğinin “sosyal karşılıklılık” alt ölçeğinden elde edilen puanların iç tutarlılığı Tablo 4.2’de yer verildiği gibi Cronbach Alfa ( $\alpha$ ) ile bulunmuştur.

**Tablo 4.2. OSBP Ölçeği ile Yapılan Puanlamalara Ait Cronbach Alfa ( $\alpha$ ) Değerleri**

	<i>Puanlayıcılar</i>				
	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>
<i>Cronbach <math>\alpha</math></i>	0.981	0.979	0.988	0.961	0.953

Tablo 4.2 incelendiğinde, her bir puanlayıcı için hesaplanan iç tutarlılık değerleri  $\alpha=0.95$  ile  $\alpha=0.98$  arasında değişmektedir. Cronbach Alfa değerlerinin 0.90’dan yüksek olması ölçeğin iç tutarlılığın oldukça yüksek olduğunu göstermektedir. Puanlamada kullanılan maddelerin Otizm Sosyal Beceriler Profili ölçeğinin sosyal karşılıklılık alt boyutunda yer alan maddeler olması yani maddelerin tek bir yapıyı ölçmesi, bu değerlerin oldukça yüksek çıkmasını açıklar niteliktedir. Elde edilen Cronbach alfa katsayılarına göre, puanlayıcıların verdiği puanların kendi içinde tutarlı olduğu yorumu yapılabilir.

##### 4.2.2. Klasik Test Kuramına Göre Otizm Sosyal Beceriler Profiline İlişkin Değerlendirmede Puanlayıcılar Arası Tutarlılık Derecesi

Beş farklı puanlayıcının aynı koşullar altında 50 bireyin her birini puanlamasına ilişkin elde edilen puanlayıcılar arası tutarlılık derecesi, parametrik olmayan istatistiksel bir teknik olan Kendall’ın uyum katsayısı (Kendall’s Concordance Coefficient) ile analiz edilmiştir. Analiz sonucunda uyum katsayısı 15 madde için

0.201 olarak bulunmuştur ( $X^2= 40.272$ ,  $sd=4$ ,  $p=0.00<0.05$ ). En az 0.80 düzeyinde olması beklenen Kendall'in uyuşum katsayısının, dereceleme ölçekleriyle gerçekleştirilen benzer çalışmalarda da (Deliceoğlu, 2009; Öztürk, 2011) beklenen düzeyden düşük hesaplandığı görülmüştür. Howell (2002)'e göre elde edilen bu sonuç doğrultusunda, puanlayıcılar arası uyum olmadığı yorumu yapılabilir. Ayrıca, 5 puanlayıcının 15 madde üzerinden verdikleri puanlar arasındaki korelasyon değerleri Tablo 4.3'de verilmiştir.

**Tablo 4.3. 5 Puanlayıcının 15 Maddeye Verdikleri Puanlar Arasındaki Korelasyon Katsayıları**

	1. <i>Puanlayıcı</i>	2. <i>Puanlayıcı</i>	3 <i>.Puanlayıcı</i>	4. <i>Puanlayıcı</i>	5. <i>Puanlayıcı</i>
<b>1.Puanlayıcı</b>	-	0.606	0.901	0.722	0.362
<b>2.Puanlayıcı</b>		-	0.578	0.727	0.585
<b>3.Puanlayıcı</b>			-	0.678	0.398
<b>4.Puanlayıcı</b>				-	0.535
<b>5.Puanlayıcı</b>					-

Tablo 4.3 incelendiğinde, 5 puanlayıcının 15 madde üzerinden verdiği puanlar arasındaki korelasyon katsayıları 0.362 ile 0.901 arasında değişmektedir. Korelasyon katsayısının, mutlak değer olarak, 0.00-0.30 arasında olması, düşük; 0.30-0.70 olması, orta; 0.70-1.00 arasında olması ise yüksek; düzeyde bir ilişki olarak tanımlanabilir (Büyüköztürk, 2002). Bu durumda, birinci ve üçüncü puanlayıcı, birinci ve dördüncü puanlayıcı son olarak da ikinci ve dördüncü puanlayıcı arasında yüksek; diğer puanlayıcılar arasında ise orta derecede ilişki vardır yorumu yapılabilir.

Ancak, Pearson çarpım moment korelasyonu, ortalamadan bağımsız olması sebebiyle, puanlayıcıların verdikleri puanlar arasındaki benzerlik ve farklılıklar hakkında bilgi verememektedir. Bu sebeple, Goodwin (2001) puanlayıcılar arası tutarlılık test edilirken, korelasyonla birlikte ortalamalarında karşılaştırılmasını önermektedir. Araştırma da korelasyon değerleri hesaplandıktan sonra, elde edilen puanların ortalamaları arasında farklılık olup olmadığı ilişkili örneklerde tek faktörlü varyans analizi ile test edilmiş ve istatistiksel olarak anlamlı bir farklılık bulunmuştur ( $F=8.261$ ,  $p=0.00<0.05$ ). Bu sonuç üzerine puanlayıcıların puanları ortalamalarının ikili karşılaştırılması için post-hoc çalışması yapılmıştır. Post-hoc çalışmasında, beşinci puanlayıcı ile diğer bütün puanlayıcıların arasında ve birinci

puanlayıcı ile dördüncü puanlayıcı arasında anlamlı bir farklılık bulunmuştur. Bu durum, Kendall'in uyum katsayısı ile elde edilen puanlayıcılar arası uyumun olmadığı yorumunu desteklemektedir.

Puanlayıcılar arası uyumun düşük çıkması sonucu, nedenlerinin araştırılması için puanlayıcılarla görüşmeler yapılmıştır. Görüşmeler sonucunda puanlayıcıların mesleki alanlarının, sosyal becerileri farklı şekilde puanlamalarına neden olduğu sonucuna varılmıştır. Örneğin, aralarında yüksek ilişki bulunan birinci puanlayıcı (psikolog) ve üçüncü puanlayıcı (sosyal psikolog, davranış terapisti) değerlendirmelere psikolog olarak yaklaştıklarını ve bunun puanlamalara yansımış olabileceğini ifade etmişlerdir.

Ayrıca, önceki bölümlerde tek tek açıklanan dereceleme ölçeklerinde güvenilirliği tehlikeye sokan dikkatsizlik, kişisel yanlılık, cömertlik/katılık, hareleme etkisi, merkeze kayma etkisi, halo etkisi gibi unsurların puanlayıcılar arası uyumu etkilemiş olabileceği düşünülmektedir.

#### **4.3. Alt problem 2: Genellenebilirlik Kuramına Göre; Birey (b), Madde (m) Ve Puanlayıcı (p) Değişkenlerinin Çapraz Tasarlandığı b x m x p Deseninin Sonuçları**

Bu bölümde, b x m x p deseninin G çalışması sonuçları ve farklı puanlayıcı ve madde sayısı senaryolarıyla oluşturulan K çalışması sonuçları iki alt başlık halinde verilmiştir.

##### **4.3.1. b x m x p Deseninin G Çalışması Sonucunda Kestirilen Varyans Bileşenleri Ve Toplam Varyansı Açıklama Yüzdeleri**

Otizm Sosyal Beceriler Profili ölçeğinin 15 maddelik "sosyal karşılıklılık" alt ölçeği 50 birey için 5 puanlayıcı tarafından puanlanmıştır. Elde edilen puanlar ile G kuramında tümüyle çaprazlanmış b x m x p (b: birey, m: madde, p: puanlayıcı) deseni uygulanmıştır. G çalışması sonucunda kestirilen varyans bileşenleri ve toplam varyansı açıklama yüzdeleri b, m ve p ana etkileri ile bm, bp, mp ve bmp ortak etkileri için Tablo 4.4'de verilmiştir.

**Tablo 4. 4. b x m x p Desenine Ait G Çalışması Sonucunda Kestirilen Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri**

<i>Varyans Kaynağı</i>	<i>Sd</i>	<i>Toplam Kareler</i>	<i>Kareler Ortalaması</i>	<i>Varyans</i>	<i>%</i>
<i>b</i>	49	1529.890	31.222	0.36541	39.9
<i>m</i>	14	76.195	5.442	0.01675	1.8
<i>p</i>	4	210.878	52.719	0.06412	7.0
<i>bm</i>	686	217.937	0.317	0.02522	2.8
<i>bp</i>	196	723.254	3.690	0.23323	25.5
<i>mp</i>	56	63.289	1.130	0.01877	2.1
<i>bmp</i>	2744	525.777	0.191	0.19161	20.9
<b>Toplam</b>	3749	3347.223			100

b:birey, m: madde, p:puanlayıcı

Tablo 4.4'te verilen G çalışması sonucunda kestirilen varyans bileşenleri ve toplam varyansı açıklama yüzdeleri incelendiğinde, en çok birey (b) ana etkisinin (%39.9), daha sonra birey x puanlayıcı (bp) ortak etkisinin (%25.5) toplam varyansı açıkladığı, en az ise %1.8 değer ile madde (m) ana etkisinin toplam varyansa katkı sağladığı görülmektedir.

Ana etkilere ait varyans bileşenleri ve toplam varyansı açıklama yüzdeleri incelendiğinde, birey (b) ana etkisinin varyansı  $\sigma_b^2(0.36541)$  olup, toplam varyansın %39.9'unu açıklamaktadır. Bu değer ile toplam varyansa en çok katkı sağlayan birey varyans bileşeni, bireylerin ölçülen özellik bakımından birbirlerinden farklılaştığını göstermektedir. Ölçme objesi olan bireylere ait varyans bileşeninin toplam varyanstaki oranının büyük olması istenilen bir durumdur (Güler, Uyanık, Teker, 2012). Bu durum, ölçme ile bireylerin bireysel farklılıklarının ortaya çıktığı ve aynı zamanda grubun heterojen bir grup olduğu şeklinde yorumlanabilir.

Madde (m) ana etkisinin varyansı  $\sigma_m^2(0.01675)$  olup, toplam varyansın %1.8'ini açıklamaktadır. Bu değer ile toplam varyansı açıklamada en son sırada yer almaktadır. Madde varyans bileşeninin toplam varyansı açıklama yüzdesinin düşük olması, her maddenin benzer güçlükte olduğunu göstermektedir.

Puanlayıcı (p) ana etkisi  $\sigma_p^2(0.06412)$  varyans değeri ile toplam varyansın %7'sini açıklamaktadır. Puanlayıcıların birbirleriyle çok benzer puanlama yapmadıkları anlamına gelen bu varyans bileşeni, ana etkilerden toplam varyansa en çok katkı sağlayan ikinci bileşendir.

Ortak etkilerden, birey x madde (bm) ortak etkisinin varyansı  $\sigma^2_{bm}$  (0.02522) olup, toplam varyansın %2.8'ini açıklamaktadır. Birey x madde etkileşiminin toplam varyansa katkısının düşük olması, bireylerin bir maddeden diğerine aldıkları puanların farklılık göstermediğini kanıtlamaktadır. Aynı zamanda bu değer, maddelerin hepsinin benzer yapıları ölçtüğü şeklinde de yorumlanabilir. Ölçekte yer alan maddelerin hepsi aynı alt ölçeğe (sosyal karşılıklılık boyutu) ait oldukları için, birey x madde etkileşiminin toplam varyansa katkısının düşük olması beklenen bir durumdur.

Birey x puanlayıcı etkileşimi (bp) ölçülmek istenmeyen puanlayıcı etkisinin, ölçülmek istenen birey etkisini etkilemesi sonucu ortaya çıkan değişkenlik kaynağıdır. Birey x puanlayıcı ortak etkisi  $\sigma^2_{bp}$  (0.23323) varyans ile toplam varyansın %25.5'ini açıklayarak, toplam varyansa getirisini en yüksek ikinci değişkenlik kaynağıdır. Bu durum göz önünde bulundurulduğunda, bireylerin bağıl durumlarının bir puanlayıcıdan diğerine değiştiği şeklinde yorum yapılabilir. Bir başka deyişle, puanlayıcılar arası tutarsızlığın olduğu görülmektedir.

Madde x puanlayıcı (mp) ortak etkisi varyansı  $\sigma^2_{mp}$  (0.01877) olup, %2.1 toplam varyansı açıklama oranıyla, toplam varyansa getirisini en düşük ikinci değişkenlik kaynağıdır. Toplam varyansı açıklama yüzdesinin düşük olması, verilen puanların bir maddeden diğerine çok farklılaşmadığını, puanlayıcıların bireyleri bir maddeden diğerine tutarlı puanladıklarını göstermektedir.

Birey x madde x puanlayıcı (bmp) etkisine ait varyans bileşeni artık varyans olarak adlandırılıp, ölçme hatasından kaynaklanmaktadır. Tablo 4.4 incelendiğinde, ölçmenin objesi olan birey (b) ve birey x puanlayıcı (bp) ortak etkileşiminden sonra toplam varyansa en büyük getirisini sağlayan üçüncü değişkenlik kaynağıdır. Artık varyans  $\sigma^2_{bmp,e}$  (0.19161) olup, toplam varyansın %20.9'unu açıklamaktadır. Artık varyans bileşeninin büyük çıkması birey, madde ve puanlayıcı ortak etkisi ve/veya tesadüfi hataların büyük olabileceğinin bir göstergesidir.

#### **4.3.2. b x m x p Deseninde Puanlayıcı ve Madde Sayılarının Artırılıp Azaltılmasıyla Yapılan Karar Çalışması Sonuçları**

Otizm Sosyal Beceriler Profili ölçeğinin "sosyal karşılıklılık" alt boyutunda yer alan 15 madde sayısı ve artırılıp azaltılarak 10, 20, 25 ve 30 olduğu, 5 puanlayıcı sayısı ve artırılıp azaltılarak 3, 4, 6 ve 7 olduğu durumlara göre düzenlenen senaryolar

için yapılan karar çalışması sonucunda kestirilen G ve Phi katsayılarına ilişkin değerler Tablo 4.5'te verilerek açıklanmıştır.

**Tablo 4.5. b x m x p Desenine Ait Karar Çalışması ile Puanlayıcı ve Madde Sayılarının Artırılıp Azaltılmasıyla Yapılan Senaryolara Göre G ve Phi Katsayıları**

Madde Sayıları	Puanlayıcı Sayıları									
	3		4		5		6		7	
	G	$\Phi$	G	$\Phi$	G	$\Phi$	G	$\Phi$	G	$\Phi$
10	0.808	0.768	0.847	0.813	0.873	0.843	0.891	0.864	0.904	0.880
15	0.813	0.774	0.852	0.819	<b>0.877</b>	<b>0.848</b>	0.895	0.869	0.908	0.885
20	0.816	0.777	0.855	0.822	0.880	0.851	0.897	0.872	0.910	0.888
25	0.817	0.779	0.856	0.823	0.881	0.853	0.898	0.874	0.911	0.889
30	0.819	0.780	0.857	0.825	0.882	0.854	0.899	0.875	0.912	0.890

Araştırmada kullanılan b x m x p deseninde 50 bireyin 5 puanlayıcı tarafından 15 madde doğrultusunda puanlanması ile elde edilen G katsayısı 0.877 olarak bulunmuştur. Tablo 4.5'deki veriler incelendiğinde, puanlayıcı sayısı sabit tutulup madde sayısı azaltıldığında ( $n_p=5$ ,  $n_m=10$ ) G katsayısı 0.873 olup, 0.004 azalırken; madde sayısı artırıldığında ( $n_p=5$ ,  $n_m=20$ ) G katsayısı 0.880 olup, 0.003 artmaktadır. Madde sayısı sabit tutulup puanlayıcı sayısı azaltıldığında ( $n_p=4$ ,  $n_m=15$ ) G katsayısı 0.852 olup, 0.025 azalırken; puanlayıcı sayısı artırıldığında ( $n_p=6$ ,  $n_m=15$ ) G katsayısı 0.895 olup, 0.018 artmaktadır. Yapılan farklı senaryolar doğrultusunda madde sayısının 5, puanlayıcı sayısının 3 olduğu durumda G katsayısı 0.808 ile en düşük değerini alırken; madde sayısının 30 puanlayıcı sayısının 7 olduğu durumda ise G katsayısı 0.912 ile en yüksek değerini almaktadır.

5 puanlayıcının 15 madde doğrultusunda yaptığı puanlamalarda Phi katsayısı ise 0.848 olarak hesaplanmıştır. Tablo 4.5'deki veriler incelendiğinde, puanlayıcı sayısı sabit tutulup madde sayısı azaltıldığında ( $n_p=5$ ,  $n_m=10$ ) Phi katsayısı 0.843 olup, 0.005 azalırken; madde sayısı artırıldığında ( $n_p=5$ ,  $n_m=20$ ) Phi katsayısı 0.851 olup, 0.003 artmaktadır. Madde sayısı sabit tutulup puanlayıcı sayısı azaltıldığında ( $n_p=4$ ,  $n_m=15$ ) Phi katsayısı 0.819 olup, 0.029 azalırken; puanlayıcı sayısı artırıldığında ( $n_p=6$ ,  $n_m=15$ ) Phi katsayısı 0.869 olup, 0.021 artmaktadır. Yapılan farklı senaryolar doğrultusunda madde sayısının 5, puanlayıcı sayısının 3

olduđu durumda Phi katsayısı 0.768 ile en düşük deęerini alırken; madde sayısının 30 puanlayıcı sayısının 7 olduđu durumda ise Phi katsayısı 0.890 ile en yüksek deęerini almaktadır.

Tüm madde ve puanlayıcı sayılarına göre G katsayısı, Phi katsayısına göre daha yüksek deęerlere sahiptir. Düzenlenen senaryolarda madde ve puanlayıcı sayısının azaltılması G ve Phi katsayılarının azalmasına, madde ve puanlayıcı sayılarının artırılması ise G ve Phi sayılarının artmasına neden olmaktadır. Tablo 4.5'deki deęerlere göre, G ve Phi katsayısını artırmada, puanlayıcı sayısını artırmanın madde sayısını artırmaya göre daha etkili olduđu görülmektedir. Elde edilen bu bulgu, Deliceođlu (2009)'un alıřmasında farklı puanlayıcı ve madde senaryolarından elde edilen G ve Phi katsayılarını artırmak için puanlayıcı sayısını artırmanın, madde sayısını artırmaya göre daha etkili olduđu řeklindeki bulgularıyla tutarlılık göstermektedir. Ayrıca, varyans bileřenleri incelendiđinde, puanlayıcı ana etkisi  $\sigma^2_p$  (0.06412) varyans deęeri ile toplam varyansın %7'sini açıklarken; madde ana etkisi varyansı  $\sigma^2_m$  (0.01675) olup, toplam varyansın %1.8'ini açıklamaktadır. Toplam varyansa getirisi bakımından, puanlayıcı ana etkisi madde ana etkisine göre daha yüksek deęere sahiptir. Bu durum, G ve Phi katsayılarını artırmada puanlayıcı sayısını artırmanın, madde sayısını artırmaya göre daha etkili olmasını destekler niteliktedir.

#### **4.4. Alt problem 3: Genellenebilirlik Kuramına Göre; Birey (b) Ve Puanlayıcı (p) Deęişkenlerinin Yuvalandıđı, Madde (m) Deęişkeninin ise apraz Tasarlandıđı (p : b) x m Deseninin Sonuları**

Bu bölümde, (p : b) x m deseninin G alıřması sonuları ve farklı puanlayıcı ve madde sayısı senaryolarıyla oluřturulan K alıřması sonuları iki alt bařlık halinde verilmiřtir.

##### **4.4.1. (p : b) x m Deseninin G alıřması Sonucunda Kestirilen Varyans Bileřenleri Ve Toplam Varyansı Açıklama Yüzdeleri**

Otizm Sosyal Beceriler Profili öleđinin "sosyal karřılıklılık" alt öleđinde yer alan 15 madde ile 50 birey 5 puanlayıcı tarafından puanlanmıřtır. Puanlayıcıların farklı bireyleri aynı maddeler dođrultusunda puanladıđı, yani bireylerin puanlayıcılarla yuvalandıđı, maddelerin ise her iki deęişkenle aprazlandıđı (p : b) x m deseni ile G alıřması yapılmıřtır. b, m, p:b, bm ve mp:b deęişkenleri için G alıřması



sonucunda kestirilen varyans bileşenleri ve toplam varyansı açıklama yüzdeleri Tablo 4.6'da verilmiştir.

**Tablo 4.6. (p : b) x m Desenine Ait G çalışması Sonucunda Kestirilen Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri**

<i>Varyans Kaynağı</i>	<i>Sd</i>	<i>Toplam Kareler</i>	<i>Kareler Ortalaması</i>	<i>Varyans</i>	<i>%</i>
<i>b</i>	49	1529.890	31.222	0.35259	39.1
<i>m</i>	14	76.195	5.442	0.02050	2.3
<i>p:b</i>	200	934.133	4.670	0.29735	33
<i>bm</i>	686	217.937	0.317	0.02146	2.4
<i>mp:b</i>	2800	589.066	0.210	0.21038	23.3
<b>Toplam</b>	<b>3749</b>	<b>3347.223</b>			<b>100</b>

b:birey, m: madde, p:puanlayıcı

Tablo 4.6'da verilen G çalışması sonucunda kestirilen varyans bileşenleri ve toplam varyansı açıklama yüzdelerine bakıldığında, en çok %39.1 ile birey (b) ana etkisinin, daha sonra %33 ile birey ve puanlayıcı (p:b) ortak etkisinin toplam varyansı açıkladığı, en az ise %2.3 değer ile madde (m) ana etkisinin toplam varyansa katkı sağladığı görülmektedir.

Ana etkilere ait varyans bileşenleri ve toplam varyansı açıklama yüzdeleri incelendiğinde, birey (b) ana etkisi  $\sigma^2_b$  (0.35259) varyans ile toplam varyansın %39.1'ini açıklayarak, toplam varyansa en çok katkı sağlayan varyans bileşenidir. Bireylere ait varyans bileşeninin toplam varyansı açıklama oranı, bireylerin ölçülen özellik bakımından farklılaştığını göstermektedir. Madde (m) ana etkisi  $\sigma^2_m$  (0.02050) varyans ile toplam varyansın %2.3'ünü açıklayarak, toplam varyansa getirisi en düşük olan bileşendir. Bu durum, ölçekte yer alan 15 maddenin zorluk-kolaylık düzeylerinin farklılık göstermediği şeklinde yorumlanabilir.

Her bir puanlayıcı farklı bireyleri puanladığı için çalışmada birey değişkeniyle puanlayıcı değişkeni yuvalanmıştır. Buradaki  $\sigma^2(b:p)$  varyans bileşeni, birey varyans bileşenini  $\sigma^2(b)$  ve birey puanlayıcı ortak etkileşim varyans bileşenini  $\sigma^2(bp)$  temsil etmektedir (Brennan, 2001). (b:p) için kestirilen varyans değeri toplam varyansın %33'ünü açıklayarak, toplam varyansa getirisi en yüksek ikinci değişkendir. Bu değer yüksek olması, birey puanlayıcı etkileşiminin farklılaştığı, bireylerin puanlarının bir puanlayıcıdan diğerine farklılık gösterdiği şeklinde

yorumlanabilir. Ortak etkilerden, birey x madde (bm) ortak etkisinin varyansı  $\sigma^2_{bm}(0.02146)$  olup, toplam varyansın %2.4'ünü açıklamaktadır. Birey x madde etkileşiminin toplam varyansa katkısının düşük olması, bireylerin bağıl durumlarının bir maddeden diğerine çok değişmediğini göstermektedir.

Tablo 4.6 incelendiğinde, ölçmenin objesi olan birey (b) ve birey puanlayıcı (p:b) etkileşiminden sonra toplam varyansa en büyük getirisi olan üçüncü değişkenlik kaynağı, artık varyans bileşenidir. Buradaki artık varyans bileşeni  $\sigma^2(mp:b,e)$ , madde puanlayıcı  $\sigma^2(mp)$  varyans bileşenini ve birey madde puanlayıcı  $\sigma^2(bmp,e)$  varyans bileşenini temsil etmektedir (Brennan, 2001). Artık varyans bileşeninin toplam varyansı açıklama yüzdesi %23.3'tür. Artık varyans bileşeninin yüksek çıkması birey madde puanlayıcı ortak etkileşimi, madde puanlayıcı ortak etkileşimi ve/veya tesadüfi hataların büyük olabileceğinin bir göstergesidir.

#### 4.4.2. (p : b) x m Deseninde Puanlayıcı ve Madde Sayılarının Artırılıp Azaltılmasıyla Yapılan Karar Çalışmaları Sonuçları

Otizm Sosyal Beceriler Profili ölçeği ile elde edilen puanlarda (p : b) x m yuvalanmış deseninde birey ölçme objesi kabul edilip, madde ve puanlayıcı sayılarının azaltılıp artırılmasıyla düzenlenen senaryolar için yapılan karar çalışması sonucunda kestirilen G ve Phi katsayılarına ait değerler Tablo 4.7'de verilerek açıklanmıştır.

**Tablo 4.7. (p : b) x m Desenine Ait Karar Çalışması ile Puanlayıcı ve Madde Sayılarının Artırılıp Azaltılmasıyla Yapılan Senaryolara Göre G ve Phi Katsayıları**

Madde Sayıları	Puanlayıcı Sayıları									
	3		4		5		6		7	
	G	$\Phi$	G	$\Phi$	G	$\Phi$	G	$\Phi$	G	$\Phi$
10	0.765	0.761	0.811	0.807	0.842	0.838	0.864	0.860	0.880	0.876
15	0.770	0.767	0.816	0.813	<b>0.846</b>	<b>0.844</b>	0.868	0.865	0.884	0.881
20	0.772	0.771	0.818	0.816	0.849	0.847	0.870	0.868	0.886	0.884
25	0.774	0.772	0.820	0.818	0.850	0.848	0.871	0.870	0.887	0.886
30	0.775	0.774	0.821	0.819	0.851	0.849	0.872	0.871	0.888	0.887

Tüm madde ve puanlayıcı sayılarına göre G katsayısı, Phi katsayısına göre daha yüksek değerlere sahiptir. Düzenlenen senaryolarda madde ve puanlayıcı sayısı

azaltıldığında G ve Phi katsayılarında azalma, madde ve puanlayıcı sayıları artırıldığında ise G ve Phi sayılarında artış görülmektedir.

Tablo 4.7’de görüldüğü üzere, 15 maddeye ve 5 puanlayıcı sayısına göre elde edilen G katsayısı 0.846 ve Phi katsayısı 0.844 olarak hesaplanmıştır. Puanlayıcı sayısı sabit tutularak madde sayısı azaltıldığında ( $n_p=5$ ,  $n_m=10$ ) G katsayısı 0.842, Phi katsayısı 0.838; madde sayısı artırıldığında ( $n_p=5$ ,  $n_m=20$ ) G katsayısı 0.849, Phi katsayısı 0.847 olarak kestirilmektedir. Madde sayısı sabit tutularak puanlayıcı sayısının azaltıldığı ( $n_p=4$ ,  $n_m=15$ ) durumda G katsayısı 0.816, Phi katsayısı 0.813; puanlayıcı sayısı artırıldığında ( $n_p=6$ ,  $n_m=15$ ) G katsayısı 0.868, Phi katsayısı 0.865 olarak kestirilmektedir. Puanlayıcı sayısı sabit tutularak madde sayısı artırıldığında ( $n_p=5$ ,  $n_m=20$ ) G katsayısında 0.03 artış görülürken, madde sayısı sabit tutulup puanlayıcı sayısı artırıldığında ( $n_p=6$ ,  $n_m=15$ ) G katsayısında 0.22 artış görülmektedir. Aynı şekilde, puanlayıcı sayısı sabit tutularak madde sayısı artırıldığından ( $n_p=5$ ,  $n_m=20$ ) Phi katsayısında 0.03 artış görülürken, madde sayısı sabit tutulup puanlayıcı sayısı artırıldığında ( $n_p=6$ ,  $n_m=15$ ) Phi katsayısında 0.021 artış görülmektedir. Buradan da görüleceği üzere, G ve Phi katsayılarını artırmada puanlayıcı sayısını artırma, madde sayısını artırmaya göre daha etkili olmaktadır. G çalışması ile elde edilen puanlayıcı varyans değerinin madde varyans değerinden yüksek olması bu durumu açıklar niteliktedir (Shavelson ve Webb, 1991).

#### **4.5. Alt Problem 4: $b \times m \times p$ ve $(p : b) \times m$ Desenlerinden Elde Edilen G Çalışması Sonuçlarının Karşılaştırılması**

Aynı birey, puanlayıcı ve madde için farklı senaryo durumlarına göre gerçekleştirilmiş  $b \times m \times p$  ve  $(p : b) \times m$  desenlerine ait G çalışması sonuçları Tablo 4.8’de verilmiştir.

**Tablo 4.8. b x m x p ve (p : b) x m Desenlerine Ait G çalışması Sonucunda Kestirilen Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri**

<i>Varyans kaynağı</i>	<i>b x m x p</i>		<i>Varyans Kaynağı</i>	<i>(p : b) x m</i>	
	<i>Varyans</i>	<i>%</i>		<i>Varyans</i>	<i>%</i>
<i>b</i>	0.36541	39.9	b	0.35259	39.1
<i>m</i>	0.01675	1.8	m	0.02050	2.3
<i>p</i>	0.06412	7.0	p:b	0.29735	33
<i>bm</i>	0.02522	2.8	bm	0.02146	2.4
<i>bp</i>	0.23323	25.5			
<i>mp</i>	0.01877	2.1			
<i>bmp</i>	0.19161	20.9	mp:b	0.21038	23.3
<i>Toplam</i>	3347.223	100		3347.223	100

Tablo 4.8'deki veriler incelendiğinde, bireylere ait varyans bileşeninin çapraz desende toplam varyansın %39.9'unu, yuvalanmış desende ise toplam varyansın %39.1'ini açıkladığı görülmektedir. Her iki desende de toplam varyansa en çok getirisini olan birey varyans bileşeni, bireylerin sosyal becerileri bakımından farklılaştığını göstermektedir. Ayrıca, çapraz desende birey ana etkisinin toplam varyansı açıklama oranının yuvalanmış desene göre daha yüksek olması, çapraz desende bireysel farklılıkların daha çok ortaya çıkarıldığı şeklinde yorumlanabilir.

Madde ana etkisinin toplam varyansı açıklama oranı çapraz desende %1.8 iken, yuvalanmış desende %2.3 olarak hesaplanmıştır. Her iki desende de madde varyans bileşeninin toplam varyansı açıklama yüzdesinin düşük olması, ölçekteki maddelerin zorluk kolaylık açısından farklılaşmadığını göstermektedir.

Çapraz desende puanlayıcı (p) varyans bileşeninin toplam varyansı açıklama oranı %7, birey x puanlayıcı (bp) ortak etkileşimine ait varyans bileşeninin ise toplam varyansı açıklama oranı %25.5'tir. Yuvalanmış desende puanlayıcı ana etkisine ve birey x puanlayıcı ortak etkisine ait varyans bileşenleri ayrı ayrı değerlendirilmek yerine, (b:p) değişkeni altında ortak değerlendirilmektedir. Aynı puanlayıcıların farklı bireyleri puanlamasıyla oluşturulan (p : b) x m deseninde bireylerin puanlayıcılarla yuvalandığı (b:p) değişkeninin varyans bileşeni toplam varyansın %33'ünü açıklamaktadır. Her iki desende de puanlayıcı varyans bileşeninin toplam varyansı açıklama oranının yüksek olduğu, dolayısıyla bireylerin durumlarının bir puanlayıcıdan diğerine değiştiği yani puanlayıcılar arası uyumun olmadığı yorumu yapılabilir.

Birey madde ortak etkisine ait varyans bileşenlerinin toplam varyansa getirisine bakıldığında, çapraz desende %2.8 iken, yuvalanmış desende %2.4 olduğu görülmektedir. Bu durum, birey puanlarının bir maddeden diğerine değişmediğini göstermektedir.

Çapraz desende madde puanlayıcı (mp) ortak etkileşimine ait varyans bileşeninin toplam varyansı açıklama oranı %2.1, birey x madde x puanlayıcı (bmp) ortak etkileşimine ait artık varyans bileşeninin toplam varyansı açıklama oranı ise %20.9'dur. Yuvalanmış desende madde puanlayıcı ortak etkisine ve artık varyansa ait varyans bileşenleri ayrı ayrı değerlendirilmek yerine, (mp:b) değişkeni altında ortak değerlendirilmektedir. Yuvalanmış desende artık varyansa ait varyans bileşeni toplam varyansın %23.3'ünü açıklamaktadır. Her iki desen içinde artık varyansın toplam varyansa getiri oranı yüksektir. Ancak bu oran, yuvalanmış desende çapraz desene göre daha yüksektir. Bu durumun yuvalanmış desende, çapraz desenden farklı olarak madde puanlayıcı ortak etkisine ait varyansın da artık varyansa dahil edilmiş olmasından kaynaklandığı yorumu yapılabilir.

#### **4.6. Alt Problem 5: b x m x p ve (p : b) x m Desenlerinde Puanlayıcı ve Madde Sayılarının Artırılıp Azaltılmasıyla Yapılan Karar Çalışmaları Sonuçlarının Karşılaştırılması**

Bu bölümde b x m x p ve (p : b) x m desenlerinde puanlayıcı ve madde sayılarının artırılıp azaltılmasıyla yapılan karar çalışmalarından elde edilen G ve Phi katsayılarının karşılaştırılmıştır. b x m x p ve (p : b) x m desenlerinde, 15 olan madde sayısının 10, 20, 25 ve 30 olduğu, 5 olan puanlayıcı sayısının ise 3, 4, 6 ve 7 olduğu senaryolar için yapılan karar çalışması sonucunda kestirilen G ve Phi katsayılarına ait değerler Tablo 4.9'da gösterilmiştir.

**Tablo 4.9. b x m x p ve (p : b) x m Desenlerine Ait Karar Çalışması ile Puanlayıcı ve Madde Sayılarının Artırılıp Azaltılmasıyla Tasarlanan Senaryolardan Elde Edilen G ve Phi Katsayıları**

M.S	D.T	Puanlayıcı Sayıları									
		3		4		5		6		7	
		G	$\Phi$	G	$\Phi$	G	$\Phi$	G	$\Phi$	G	$\Phi$
10	Ç.D	0.808	0.768	0.847	0.813	0.873	0.843	0.891	0.864	0.904	0.880
	Y.D	0.765	0.761	0.811	0.807	0.842	0.838	0.864	0.860	0.880	0.876
15	Ç.D	0.813	0.774	0.852	0.819	<b>0.877</b>	<b>0.848</b>	0.895	0.869	0.908	0.885
	Y.D	0.770	0.767	0.816	0.813	<b>0.846</b>	<b>0.844</b>	0.868	0.865	0.884	0.881
20	Ç.D	0.816	0.777	0.855	0.822	0.880	0.851	0.897	0.872	0.910	0.888
	Y.D	0.772	0.771	0.818	0.816	0.849	0.847	0.870	0.868	0.886	0.884
25	Ç.D	0.817	0.779	0.856	0.823	0.881	0.853	0.898	0.874	0.911	0.889
	Y.D	0.774	0.772	0.820	0.818	0.850	0.848	0.871	0.870	0.887	0.886
30	Ç.D	0.819	0.780	0.857	0.825	0.882	0.854	0.899	0.875	0.912	0.890
	Y.D	0.775	0.774	0.821	0.819	0.851	0.849	0.872	0.871	0.888	0.887

M.S: Madde Sayıları, D.T: Desen Türü, Ç.D: Çapraz Desen, Y.D: Yuvalanmış Desen

50 bireyden her birinin, 5 puanlayıcı tarafından 15 madde doğrultusunda puanlanmasıyla oluşturulan tümüyle çaprazlanmış b x m x p deseninde G katsayısı 0.877, Phi katsayısı ise 0.848 olarak hesaplanmıştır. Aynı verilerle gerçekleştirilen, her bir puanlayıcının bu kez farklı öğrencileri aynı maddeler doğrultusunda puanladığı yani bireylerin puanlayıcılarla yuvalandığı, maddelerin ise her iki değişkenle çaprazlandığı (p : b) x m deseninde G katsayısı 0.846, Phi katsayısı ise 0.844 olarak kestirilmiştir. b x m x p deseni ile kestirilen G katsayısı, (p : b) x m deseni ile yapılan karar çalışması sonucu elde edilen G katsayısından 0.031 daha büyük hesaplanmıştır. Aynı şekilde, b x m x p deseni ile kestirilen Phi katsayısı, (p : b) x m deseni ile yapılan karar çalışması sonucu elde edilen Phi katsayısından 0.04 daha büyük bulunmuştur. Böylelikle; her iki desenden tümüyle çaprazlanmış olan b x m x p deseninde G ve Phi katsayılarının daha yüksek hesaplandığı görülmektedir.

b x m x p deseninde puanlayıcı sayısı sabit tutulup madde sayısı azaltıldığında ( $n_p=5$ ,  $n_m=10$ ) G katsayısı 0.873, madde sayısı artırıldığında ( $n_p=5$ ,  $n_m=20$ ) G katsayısı 0.880 olmaktadır. (p : b) x m deseninde ise puanlayıcı sayısı sabit tutulup madde sayısının azaltıldığı durumda ( $n_p=5$ ,  $n_m=10$ ) G katsayısı 0.842

hesaplanırken, madde sayısının artırıldığı durumda ( $n_p=5$ ,  $n_m=20$ ) G katsayısı 0.849 olarak hesaplanmaktadır. Aynı durum Phi katsayıları için incelendiğinde,  $b \times m \times p$  deseninde puanlayıcı sayısı sabit tutulup madde sayısı azaltıldığında ( $n_p=5$ ,  $n_m=10$ ) Phi katsayısı 0.843, madde sayısı artırıldığında ( $n_p=5$ ,  $n_m=20$ ) Phi katsayısı 0.851 olmaktadır.  $(p : b) \times m$  deseninde ise puanlayıcı sayısı sabit tutulup madde sayısının azaltıldığı durumda ( $n_p=5$ ,  $n_m=10$ ) Phi katsayısı 0.838 hesaplanırken, madde sayısının artırıldığı durumda ( $n_p=5$ ,  $n_m=20$ ) Phi katsayısı 0.847 olarak hesaplanmaktadır. Bu değerlere göre, puanlayıcı sayısının sabit tutulup, madde sayısının azaltılıp artırıldığı durumlarda  $b \times m \times p$  deseninde kestirilen G ve Phi katsayılarının  $(p : b) \times m$  deseninde kestirilen G ve Phi katsayılarından daha yüksek olduğu görülmektedir.

Tablo 4.9'daki verilere göre, bu kez çapraz desende madde sayısının sabit tutularak puanlayıcı sayısının azaltıldığı durumda ( $n_p=4$ ,  $n_m=15$ ) G katsayısının 0.852; puanlayıcı sayısının artırıldığı durumda ise ( $n_p=6$ ,  $n_m=15$ ) G katsayısının 0.895 olduğu görülmektedir. Yuvalanmış desende ise madde sayısı sabit tutularak puanlayıcı sayısı azaltıldığında ( $n_p=4$ ,  $n_m=15$ ) G katsayısının 0.816; puanlayıcı sayısı artırıldığında ( $n_p=6$ ,  $n_m=15$ ) ise G katsayısının 0.868 olduğu görülmektedir. Aynı senaryolar için Phi katsayısı incelenirse,  $b \times m \times p$  deseninde madde sayısı sabit tutulup puanlayıcı sayısı azaltıldığında ( $n_p=4$ ,  $n_m=15$ ) Phi katsayısı 0.819 olarak kestirilirken; puanlayıcı sayısı artırıldığında ( $n_p=6$ ,  $n_m=15$ ) ise Phi katsayısı 0.869 olarak kestirilmektedir. Bu durum  $(p : b) \times m$  deseninde, madde sayısı sabit tutularak puanlayıcı sayısının azaltıldığı ( $n_p=4$ ,  $n_m=15$ ) durumda Phi katsayısı 0.813; puanlayıcı sayısı artırıldığında ( $n_p=6$ ,  $n_m=15$ ) ise Phi katsayısı 0.865 olarak kestirilmektedir. Böylece, puanlayıcı sayısının sabit tutulup, madde sayısının azaltılıp artırıldığı durumlarda çapraz desende kestirilen G ve Phi katsayılarının yuvalanmış desende kestirilen G ve Phi katsayılarından daha yüksek olduğu görülmektedir.

Shavelson ve Webb (1991)'e göre "yüksek" güvenilirlik sağlayabilmek için G ve Phi katsayılarının en az 0.80 olması gerektiğini belirtmişlerdir. Bu durumun sağlanabilmesi için her iki desende de puanlayıcı sayısının en az 4 olması gerekmektedir. Bu durumda, zaman ve işgücü göz önünde bulundurularak yüksek güvenilirlikli en ekonomik uygulamanın 4 puanlayıcı ve 10 madde ile yapılacağı söylenebilir. 4 puanlayıcı ve 10 madde için  $b \times m \times p$  deseninde elde edilen G

katsayısı 0.847 ve Phi katsayısı 0.813 olarak kestirilirken; (b : p) x p deseninde elde edilen G katsayısı 0.811 ve Phi katsayısı 0.807 olarak kestirilmektedir.

Genel olarak Tablo 4.9 incelendiğinde, her iki desende de düzenlenen tüm senaryolarda madde ve puanlayıcı sayısı azaltıldığında G ve Phi katsayılarında azalma, madde ve puanlayıcı sayıları artırıldığında ise G ve Phi sayılarında artış görülmektedir. Düzenlenen tüm senaryolarda çapraz desende kestirilen G ve Phi katsayılarının yuvalanmış desende kestirilen G ve Phi katsayılarından daha yüksek olduğu görülmektedir. Ayrıca, gerek çapraz desende gerekse yuvalanmış desende tüm senaryo durumlarına göre G katsayıları Phi katsayılarından daha yüksek kestirilmektedir. Benzer çalışmalarda (Nalbantoğlu, 2009; Alkan, 2013) da G katsayılarının Phi katsayılarından yüksek kestirildiği görülmektedir. Bu durum, mutlak ve bağıl hata varyansları arasındaki farklılıktan kaynaklandığı için beklenen bir durumdur. Tanımı gereği mutlak hata varyansı, bağıl hata varyansından daha yüksek bir değere sahip olduğu için, Phi katsayısı G katsayısına göre daha düşük bir değere sahip olur.



## 5. SONUÇ ve ÖNERİLER

Bu bölümde araştırmanın bulgularına dayalı olarak ulaşılan sonuçların özetine ve bu sonuçlar doğrultusunda yapılabilecek önerilere yer verilmiştir.

### 5.1. Sonuçlar

#### 5.1.1. Birinci Alt Probleme İlişkin Sonuçlar

1) Otizm Sosyal Beceriler Profili ölçeğinin “sosyal karşılıklılık” alt boyutunun klasik test kuramına göre iç tutarlılık düzeyinin belirlenmesi için Cronbach  $\alpha$  güvenilirlik katsayısı hesaplanmıştır. 5 puanlayıcı için ayrı ayrı hesaplanan Cronbach  $\alpha$  değerleri oldukça yüksek çıkmıştır. Bu doğrultuda, ölçekte yer alan maddelerin birbirleriyle oldukça tutarlı olduğu, aynı zamanda her bir puanlayıcının kendi içinde tutarlı puanlamalar yaptığı sonucu çıkarılmıştır.

2) Otizm Sosyal Beceriler Profili ölçeğinin puanlanmasında yer alan 5 puanlayıcı arasındaki uyum düzeyinin belirlenmesi için klasik test kuramında ilk olarak Kendall’ın uyum katsayısı hesaplanmıştır. Ayrıca, her bir puanlayıcının verdiği puanlar ile diğer puanlayıcıların verdikleri puanlar arasındaki ilişkiler Pearson momentler çarpım korelasyon katsayısı ile hesaplanmıştır. Daha sonra, puanlayıcıların verdikleri puanların ortalamaları arasında manidar bir farklılığın olup olmadığının belirlenmesi için ilişkili örneklemelerde tek faktörlü varyans analizi kullanılmıştır. Yapılan analizler sonucunda, puanlayıcıların katılık/cömertlik düzeylerinin birbirlerinden farklı olduğu, puanlayıcıların puanlamalarının birbirleriyle paralellik göstermediği ve bu sebeplerle puanlayıcılar arası tutarsızlığın olduğu görülmüştür.

#### 5.1.2. İkinci Alt Probleme İlişkin Sonuçlar

1) OSBP ile elde edilen puanlarla Genellenebilirlik Kuramında birey, madde ve puanlayıcı değişkenlerinin çaprazlandığı  $b \times m \times p$  deseninde G çalışması yapılmış, kestirilen varyans bileşenleri ve toplam varyansı açıklama yüzdeleri bulunmuştur.

a) G çalışmasına ait bulgular incelendiğinde, toplam varyansı açıklama yüzdesi en yüksek olan değişkenlik kaynağın birey (b) ana etkisi olduğu görülmüştür. Birey ana etkisinin, ölçmenin objesi olması nedeniyle bu durum istenilen bir durumdur.

Buna göre, OSBP ile gerçekleştirilen ölçme işlemi sonucunda bireyler arası farklılıkların ortaya çıkarılabildiği sonucuna varılmıştır.

b) Madde ana etkisi (m), birey madde ortak etkisi (bm) ve madde puanlayıcı (mp) ortak etkisine ait varyansların düşük olması; sırasıyla ölçekte yer alan maddelerin zorluk/kolaylık düzeylerinin benzer olduğunu, bireylerin bir maddeden diğerine aldıkları puanların farklılık göstermediğini ve puanlayıcıların verdikleri puanların bir maddeden diğerine değişmediğini göstermiştir.

c) Birey puanlayıcı (bp) ortak etkisine ait varyans bileşeni, toplam varyansa katkısı en yüksek olan ikinci değişkenlik kaynağıdır. Bu değerlerin yüksek olması, bireylerin bağıl durumlarının bir puanlayıcıdan diğerine değiştiğini göstermiştir. Bu nedenle puanlayıcılar arası uyumun düşük olduğu yorumu yapılabilir. Puanlayıcı (p) değişkenine ait varyans bileşeninin, ana etkiler arasında ölçme objesinden sonra en yüksek varyansa sahip bileşen olması, bu durumu destekler niteliktedir.

d) Birey madde puanlayıcı (bmp) ortak etkisine ait varyans bileşeni, yani artık varyans, toplam varyansa getirisi en yüksek olan üçüncü varyans bileşenidir. Artık varyans bileşeninin büyük çıkması araştırmada; birey, madde ve puanlayıcı ortak etkisi ve/veya tesadüfi hataların büyük olabileceğinin bir göstergesidir.

2) Tümüyle çaprazlanmış  $b \times m \times p$  deseninde, Otizm Sosyal Beceriler Profili ölçeğinin “sosyal karşılıklık” alt boyutunda yer alan 15 madde sayısına ve bu maddeleri puanlayan 5 puanlayıcıya göre elde edilen G katsayısı (0.877) ve Phi katsayısı (0.848) birbirine yakın ve yüksek değerler olarak elde edilmiştir. Madde sayısının sabit tutularak, puanlayıcı sayısının azaltılıp artırıldığı ve puanlayıcı sayısının sabit tutularak, madde sayısının azaltılıp artırıldığı analizler sonucunda, G ve Phi katsayılarını yükseltmek için puanlayıcı sayısını artırmanın, madde sayısını artırmaya göre daha etkili olduğu sonucuna varılmıştır. G çalışmasında elde edilen puanlayıcı varyans bileşeninin toplam varyansı açıklama yüzdesinin, madde varyans bileşeninin toplam varyansı açıklama yüzdesine oranla daha yüksek olması bu sonucu destekler niteliktedir.

### 5.1.3. Üçüncü Alt Probleme İlişkin Sonuçlar

1) Aynı verilerle Genellenebilirlik kuramında bu kez birey ve puanlayıcı değişkenlerinin birbiriyle yuvalandığı, madde değişkeninin ise her iki değişkenle

çaprazlandığı (p : b) x m deseninde G çalışması yapılmış, kestirilen varyans bileşenleri ve toplam varyansı açıklama yüzdeleri bulunmuştur.

a) G çalışması sonucunda toplam varyansa katkı sağlayan 5 varyans bileşeninden, yüzdesi en yüksek olan birey (b) değişkenine ait varyans bileşenidir. Bireylerin ölçme objesi olduğu göz önüne alınınca yüksek olması beklenen birey değişkenine ait toplam varyansı açıklama yüzdesi, bireylerin ölçülen özellik bakımından birbirinden ayrıldığını göstermiştir.

b) Madde ana etkisine ait varyans değerinin toplam varyansa getirisinin düşük çıkması, maddelerin güçlük düzeylerinin birbirine yakın olduğunu göstermiştir. Madde puanlayıcı ortak etkisine ait varyans değerinin toplam varyansa getirisinin düşük olması ise birey puanlarının maddeden maddeye değişmediği şeklinde yorumlanmıştır.

c) Birey değişkeniyle puanlayıcı değişkeninin yuvalandığı (p:b) değişkenine ait varyans değeri, toplam varyansa en çok katkıyı sağlayan ikinci bileşendir. Bu değer yüksek çıkması, birey puanlayıcı etkileşiminin farklılaştığını yani puanlayıcılar arası tutarsızlık olduğunu göstermiştir.

d) K çalışması sonucunda artık varyans değeri ise oldukça yüksek çıkmıştır. Bu değer yüksek olması, birey madde puanlayıcı ortak etkileşimi, madde puanlayıcı ortak etkileşimi ve/veya tesadüfi hataların büyük olabileceğinin bir göstergesidir.

2) Birinci desende kullanılan aynı verilerle oluşturulan (p : b) x m deseninin karar çalışmasında madde ve puanlayıcı sayılarında yapılan değişikliklere göre G ve Phi katsayıları hesaplanmıştır. 5 puanlayıcınının 50 öğrenciyi dönüşümlü olarak 15 madde doğrultunda puanlaması sonucu elde edilen G katsayısı 0.846 ve Phi katsayısı 0.844 olarak hesaplanmıştır. Madde sayısının sabit tutulup puanlayıcı sayısının azaltılarak 2 ve 4 yapıldığı durumlarda G ve Phi katsayılarında düşüş; puanlayıcı sayılarının artırılarak 6 ve 7 olduğu durumlarda G ve Phi katsayılarında artış meydana gelmiştir. Aynı şekilde puanlayıcı sayısının sabit tutulup madde sayısının azaltılarak 10 yapıldığı durumda G ve Phi katsayılarında düşüş; madde sayılarının artırılarak 20, 25 ve 30 olduğu durumlarda G ve Phi katsayılarında artış meydana gelmiştir. Ancak puanlayıcı varyans değerinin madde varyans değerinden yüksek olması sebebiyle, G ve Phi katsayıları puanlayıcı sayısındaki değişikliklerden daha çok etkilenmişlerdir.

#### 5.1.4. Dördüncü Alt Probleme İlişkin Sonuçlar

$b \times m \times p$  çapraz ve  $(p : b) \times m$  yuvalanmış desenlerinde yapılan G çalışmalarıyla elde edilen varyans ve toplam varyansı açıklama yüzdeleri karşılaştırılmıştır.

a) Ölçme objesi olan bireylere ait varyans değeri, her iki desende de toplam varyansa en çok getirisini sağlayan bileşen olmuştur.  $b \times m \times p$  deseninde kestirilen madde varyans bileşeninin toplam varyansı açıklama yüzdesi,  $(p : b) \times m$  deseninde kestirilen yüzdeye göre daha yüksek olsa da, her iki desende de bireylerin sosyal becerileri bakımından farklılaştığı görülmüştür.

b) Her iki desende de madde ana etkisine ait varyans değerlerinin toplam varyansı açıklama oranlarının düşük olması nedeniyle, maddelerin güçlük düzeylerinin farklılaşmadığı sonucu çıkarılmıştır.

c) Çapraz desende puanlayıcı ana etkisi ve birey puanlayıcı ortak etkisi ayrı ayrı hesaplanırken, yuvalanmış desende bu iki değişkenlik kaynağı  $(p:b)$  bileşeni altında hesaplanmıştır. Her iki desende de puanlayıcılarla ilgili varyans bileşenlerinin yüksek olduğu görülmüş, buna dayalı olarak da puanlayıcılar arası uyumun düşük olduğu sonucu çıkarılmıştır.

d) Birey madde ortak etkisine ait varyans bileşenlerine bakıldığında, her iki desende de toplam varyansa getirilerinin düşük olduğu ve aralarında farklılık olmadığı görülmüştür. Buna dayalı olarak, bireylerin puanlarının maddeden maddeye değişmediği sonucu çıkarılmıştır.

e)  $b \times m \times p$  deseninde madde puanlayıcı ortak etkisine ait varyans ve artık varyans ayrı ayrı hesaplanırken,  $(p : b) \times m$  deseninde  $(mp:b)$  değişkeni altında ortak hesaplandığı görülmüştür. Her iki desende de artık varyansın toplam varyansı açıklama yüzdesi yüksek çıkmış olsa da,  $(p : b) \times m$  deseninde artık varyans daha yüksek bulunmuştur. Bunun sebebinin çapraz desenden farklı olarak,  $(p : b) \times m$  deseninde madde puanlayıcı ortak etkisine ait varyans değerinin de artık varyansa dahil edilmesinden kaynaklandığı yorumu yapılabilir.

Genel olarak, çapraz ve yuvalanmış desenden elde edilen bulgulara göre kestirilen varyans değerleri arasında paralellik olduğu ancak, çapraz desenin birey puanlayıcı ve madde puanlayıcı ortak etkileri içinde bilgi üretmesi sebebiyle daha detaylı bilgi verdiği yorumu yapılabilir.

### **5.1.5. Beşinci Alt Probleme İlişkin Sonuçlar**

$b \times m \times p$  ve  $(p : b) \times m$  desenlerinde yapılan karar çalışmalarında puanlayıcı ve madde sayılarının artırılıp azaltılmasıyla elde edilen G ve Phi katsayıları karşılaştırıldığında, madde ve puanlayıcı sayısının azaltıldığı durumlarda G ve Phi katsayılarında azalma, madde ve puanlayıcı sayılarının artırıldığı durumlarda ise G ve Phi sayılarında artış görülmüştür. Düzenlenen tüm senaryolarda çapraz desende kestirilen G ve Phi katsayıları, yuvalanmış desende kestirilen G ve Phi katsayılarından daha yüksek kestirilmiştir. Her iki desen içinde kabul edilebilir 0.80 sınırında G ve Phi katsayısı elde edebilmek için puanlayıcı sayısının en az 4 olması gerektiği belirlenmiştir.

## **5.2. Öneriler**

### **5.2.1. Araştırmadan Çıkarılan Öneriler**

- 1) Dereceleme ölçeklerinin kullanıldığı birden fazla puanlayıcının değerlendirme yaptığı ölçümlerde, asıl ölçek kullanılmadan önce belli bir düzeyde puanlayıcılar arası güvenilirlik elde edebilmek için başka ölçeklerle ön çalışmalar yapılabilir.
- 2) Çalışma sonucunda elde edilen puanlayıcılar arası uyumun düşük olmasının nedenleri nitel çalışmalar ile araştırılabilir.
- 3) Otizm Sosyal Beceriler Profili ölçeği ile yapılan çalışmalarda 4 puanlayıcı kullanılarak kabul edilebilir düzeyde güvenilirlik sağlandığı için çalışmalarda 4 puanlayıcının kullanılması zaman, ekonomi ve emek açısından daha uygun olabilir.

### **5.2.2. Yapılacak Başka Araştırmalar için Öneriler**

- 1) Birden fazla puanlayıcının puanlama yapması gereken değerlendirmelerde, puanlayıcılar arası güvenilirlik test edilirken Klasik Test Kuramına göre Kendall'ın uyum katsayısı toplam puanlar üzerinden analiz yaparken, Genellenebilirlik Kuramı maddeler bazında analiz yaptığı için daha detaylı bilgi vermektedir. Bu sebeple, puanlayıcılar arası güvenilirlik analizlerinde genellenebilirlik kuramı tercih edilebilir.
- 2) KTK güvenilirlik hesaplamasında sadece bağıl değerlendirmeler için güvenilirlik hesaplanırken, GK hem bağıl değerlendirme hem de mutlak değerlendirmeler için güvenilirlik katsayıları üretir. Bu sebeple mutlak değerlendirmelerin söz konusu olduğu durumlarda GK tercih edilebilir.

3) Bu alıřmada Otizm Sosyal Beceriler Profili leđinin gvenirliđi Genellenebilirlik Kuramının  $b \times m \times p$  ve  $(p : b) \times m$  modelleri ile analiz edilmiřtir. Benzer alıřmalar GK'da farklı modeller ve farklı deđiřkenlik kaynakları ele alınarak yapılabilir.

4) Klasik Test Kuramında puanlayıcılar arası tutarlılıđı belirlemek amacıyla farklı teknikler ile bu teknikleri birbirleriyle kıyaslayan arařtırmalar yapılabilir.

5) Farklı trde lmlerde KTK, GK ve Madde Tepki Kuramını birbirleriyle kıyaslayan alıřmalar yapılabilir.

## KAYNAKÇA

- Alkan, M. (2013). *PISA 2009 okuma becerileri açık uçlu sorularının puanlanmasında genellenebilirlik kuramındaki farklı desenlerin karşılaştırılması*. Yayınlanmamış doktora tezi. Hacettepe Üniversitesi, Ankara.
- Asher, S., Singleton, L., Tinsley, B., & Hymel, S. (1979). A reliable sociometric measure for preschool children. *Developmental Psychology*, 15(4), 443-444.
- Atılğan, H. (2004). *Genellenebilirlik kuramı ve çok değişkenlik kaynaklı Rasch modelinin karşılaştırılmasına ilişkin bir araştırma*. Yayınlanmamış doktora tezi. Hacettepe Üniversitesi, Ankara.
- Atılğan, H. (2008). Using generalizability theory to assess the score reliability of the special ability selection examinations for music education programmes in higher education. *International Journal of Research and Method Education*, 31(1), 63-76.
- Atılğan, H. ve Tezbaşaran, A. A. (2005). Genellenebilirlik kuramı alternatif karar çalışmaları ile senaryolar ve gerçek durumlar için elde edilen G ve Phi katsayılarının tutarlılığının incelenmesi. *Eğitim Araştırmaları Dergisi*, 18, 28-40.
- Au, F., Prahardhi, S., & Shiell, A. (2008). Reliability of two instruments for critical assessment of economic evaluations. *Value in Health*, 11(3), 435-439.
- Avcioğlu, H. (2001). *İşitme engelli çocuklara sosyal becerilerin öğretilmesinde işbirlikçi öğrenme yaklaşımı ile sunulan öğretim programının etkililiğinin incelenmesi*. Yayınlanmamış doktora tezi. Hacettepe Üniversitesi, Ankara.
- Avcioğlu, H. (2011). *Özel gereksinimi olan bireylerin değerlendirilmesi*. Ankara: Vize Yayıncılık.
- Balcı, A. (2001). *Sosyal bilimlerde araştırma: yöntem, yeti ve ilkeler*. Ankara: Pegem Akademi.
- Bauminger, N., Shulman, C., & Agam, G. (2003). Peer interaction and loneliness in high-functioning children with autism. *Journal of Autism and Developmental Disorders*, 33(5), 489-507.
- Turgut, M. F. ve Baykul, Y. (2014). *Eğitimde ölçme ve değerlendirme*. (6. Baskı). Ankara: Pegem Akademi.
- Baykul, Y. (2000). *Eğitimde ve psikolojide ölçme: klasik test teorisi ve uygulaması*. Ankara: ÖSYM Yayınları.
- Bellini, S., & Hopf, A. (2007). The development of the autism social skills profile: A preliminary analysis of psychometric properties. *Focus on Autism and Other Developmental Disabilities*, 22(2), 80-87.
- Brennan, R. L. (2000). Performance assessments from the perspective of generalizability theory. *Applied Psychological Measurement*, 24(4), 339-353.
- Brennan, R. L. (2001). *Generalizability theory*. New York: Springer-Verlog.

- Bursuck, W. D., & Asher, S. R. (1986). The relationship between social competence and achievement in elementary school children. *Journal of Clinical Psychology*, 15, 41–49.
- Büyükkıdık, S. (2012). *Problem çözme becerisinin değerlendirilmesinde puanlayıcılar arası güvenilirliğin klasik test kuramı ve genellenebilirlik kuramına göre karşılaştırılması*. Yayınlanmamış yüksek lisans tezi. Hacettepe Üniversitesi, Ankara.
- Büyükoztürk, Ş. (2013). *Sosyal bilimler için veri analizi el kitabı*. (18. Baskı). Ankara: Pegem Akademi.
- Cardinet, J., Johnson, S., & Pini, G. (2009). *Applying generalizability theory using eduG (Quantitative methodology series)*. New York, London: Routledge.
- Chen, D., Hu, B. Y., Fan, X., & Li, K. (2014). Measurement quality of the Chinese Early Childhood Program Rating Scale: An investigation using multivariate generalizability theory. *Journal of Psychoeducational Assessment*, 32(3), 236-248.
- Choi, H. S., & Heckenlaible-Gotto, M. J. (1998). Classroom-based social skill training: Impact on peer acceptance of first-grade students. *Journal of Educational Research*, 91, 209–214.
- Christ, T. J., Tillman C., Chafouleas, S. M., & Boice C. H. (2010). Direct Behavior Rating (DBR): Generalizability and dependability across raters and observations. *Educational and Psychological Measurement*, 70(5), 825–843.
- Cohen, J. R., Swerdlik, E. M., & Phillips, S. M. (1996). *Psychological testing and assesment*. (7th edition). London: Mayfield Publishing Company.
- Crocker, L. M., & Algina, L. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston.
- Crozier, S., & Tincani, M. (2005). Using a modified social story to decrease disruptive behavior of a child with autism. *Focus on Autism and Other Developmental Disabilities*, 20(3), 150-157.
- Deliceoğlu, G. (2009). *Futbol yetilerine ilişkin dereceleme ölçeğinin genellenebilirlik ve klasik test kuramına dayalı güvenilirliklerinin karşılaştırılması*. Yayınlanmamış doktora tezi. Ankara Üniversitesi, Ankara.
- Demir, Ş. (2009). *Otizmli çocukların sosyal becerilerinin farklı değişkenler açısından değerlendirilmesi*. Yayınlanmamış yüksek lisans tezi. Ankara Üniversitesi, Ankara.
- Demir, Ş. (2012). Otizm spektrum bozukluğu olan çocuklara sosyal becerilerin öğretimi. Elif Tekin-İftar (Ed). *Otizm spektrum bozukluğu olan çocuklar ve eğitimleri*, 369-423. Ankara: Vize Yayıncılık
- Doğan, N. (2002). *Klasik test teorisi ve örtük özellikler kuramının örneklemeler bağlamında karşılaştırılması*. Yayınlanmamış doktora tezi. Hacettepe Üniversitesi, Ankara.



- Eason, S. H. (1989). *Why generalizability theory yields beter results than classical test theory*. Mid-South Educational Research Association Annual Meeting, 8-10 November 1989.
- Elliott, S. N., & Gresham, F. M. (1987). Children's social skills: assessment and classification practices. *Journal of Counseling & Development*, 66(2), 96-99.
- Eser, D. ve Gelbal, S. (2013). Genellenebilirlik kuramı ve lojistik regresyona dayalı hesaplanan puanlayıcılar arası tutarlılığın karşılaştırılması. *Kastamonu Eğitim Dergisi*, 21(2), 421-438.
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5), 378-382
- Goodwin, L. D. (2001). Interrater agreement and reliability. *Measurement in Psychical Education and Exercises Science*, 5(1), 13-34.
- Gresham, F. M. (1986). Conceptual and definitional issues in the assesment of childrens's social skills: implications for classification and training. *Journal of Clinical Child Psychology*, 15(1), 3-15.
- Güler, N. ve Gelbal S. (2010). Açık uçlu matematik sorularının güvenilirliğinin klasik test kuramı ve genellenebilirlik kuramına göre incelenmesi. *Kuram ve Uygulamada Eğitim Bilimleri Dergisi*, 10(2), 989–1019.
- Güler, N. (2008). *Klasik test kuramı genellenebilirlik kuramı ve Rasch modeli üzerine bir araştırma*. Yayınlanmamış doktora tezi. Hacettepe Üniversitesi, Ankara.
- Güler, N. (2011). Rasgele veriler üzerinde genellenebilirlik kuramı ve klasik test kuramına göre güvenirlğin karşılaştırılması. *Eğitim ve Bilim*, 36(162), 225-234.
- Güler, N., Uyanık, G. K. ve Teker, G. T. (2012). *Genellenebilirlik Kuramı*. Ankara: Pegem Akademi
- Güler, N., Eroğlu, Y. ve Akbaba, S. (2014). Reliability of criterion-dependent measurement tools according to generalizability theory: application in the case of eating skills. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 14(2), 217-232.
- Hall, J. A., Schlesinger, D. J., & Dieen, J. P. (1997). Social skills training in group with developmentally disabled adults. *Research on Social Work Practice*, 7(2), 187-201.
- Hambelton, R. K., & Swaminathan, H. (1985). *Item response theory: principles and application*. Kluwer: Nijhoff Publishing.
- Hartmann, D. P. (1977). Considerations in the choice of interobserver reliability estimates. *Journal of Applied Behavior Analysis*, 10, 103–116.
- Heward, W. L. (2012). *Exceptional Children. An introduction to special education*. (10th edition). London: Pearson.
- Howell, D. C. (2009). *Statistical methods for psychology*. (7th Edition). USA: Thomson Learning Academic Research Center.

- Kantor, R., & Lee, Y. W. (2007). Evaluating prototype tasks and alternative rating schemes for a new ESL writing test through G theory. *International Journal of Testing*, 7(4), 353-385.
- Kırcaali-İftar, G. (2012). Otizm spektrum bozukluğuna genel bakış. E. Tekin-İftar (Ed). *Otizm spektrum bozukluğu olan çocuklar ve eğitimleri*, 17-46. Ankara: Vize Yayıncılık.
- Kıroğlu, G. (2001). *Uygulamalı parametrik olmayan istatistiksel yöntemler*. İstanbul: Paymaş Yayınları.
- Kieffer, K. M. (1998). *Why generalizability theory is essential and classical test theory is often inadequate?* Paper presented at the annual meeting of the Southwestern Psychological Association. New Orleans, LA. USA.
- La Grace, A. M. (1993). Social skills training with children: Where do we go from here? *Journal of Clinical Child Psychology*, 22(1), 288-298.
- Landis, R. J., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.
- Lane, K. L., Givner, C. C., & Pierson, M. R. (2004). Teacher expectations of student behavior: Social skills necessary for success in elementary school classrooms. *The Journal of Special Education*, 38(2), 104-110.
- Lee, G., & Fitzpatrick, A. R. (2003). The effects of a student sampling plan on estimates of the standard errors for student passing rates. *Journal of Educational Measurement*, 40(1), 17-28.
- Lee, G., & Frisbie, D. A., (1999). Estimating reliability under a generalizability theory model for test scores composed of testlets. *Applied Measurement in Education*, 12(3), 237-255.
- Lee, Y. W. (2005). Dependability of scores for a new ESL speaking test: evaluating prototype tasks. Monograph Series MS-28. Princeton, NJ: *Educational Testing Service*.
- Liberman, R.P., Mueser, K.T., Wallace, C.J., Jacobs, H.E., Eckman, T., & Massel H.K. (1986). Training skills in the psychiatrically disabled: Learning Coping And Competence. *Schizophrenia Bulletin*, 12, 631-647.
- Lorber, M. F. (2006). Can minimally trained observers provide valid global ratings? *Journal of Family Psychology*, 20(2), 335-338.
- Lord, F. M., & Novick, R. M. (1968). *Statistical theories of mental test scores*. California: Addison-Wesley Publishing Company.
- MacMillian, P. D. (2000). Classical, generalizability and multifaceted rasch detection interrater variability in large, sparse data set. *Journal of Experimental Education*, 68(2), 67-194.
- Merrell, K. W., & Carderalla, P. (1999). Social-behavioral assesment of at-risk early adolescent students: psychometric characteristics and validity of a parent report form of the school social behavior scales. *Journal Of Psychoeducational Assesment*, 17, 36-49.

- Merrell, K.W. (2001). Assessments of children's social skills: recent developments, best practices and new directions. *Exceptionality*, 9(1&2), 3-18.
- Mushquash, C., & O'Connor, B. P. (2006). SPSS And SAS programs for generalizability theory analyses. *Behavior Research Methods*, 38(3), 542-547.
- Nalbantođlu, F. (2009). *Performans ölçümlerinde genellenabilirlik kuramıyla farklı desenlerin karşılaştırılması*. Yayımlanmamış yüksek lisans tezi. Hacettepe Üniversitesi, Ankara.
- Özçelik, D. A. (1991). *Ölçme ve değerlendirme*. (2. Baskı). Ankara: ÖSYM Yayınları
- Özdemir, O., Diken, İ. H., Diken, Ö. ve Şekerciođlu, G. (2013). Otizm Davranış Kontrol Listesi (Autism Behavior Checklist-ABC) modifiye edilmiş Türkçe versiyonunun geçerlik ve güvenilirlik çalışması: pilot uygulama sonuçları. *International Journal of Early Childhood Special Education (INT-JECSE)*, 5(2), 168-186.
- Öztürk, M. E. (2011). *Voleybol Becerileri Gözlem Formu ile elde edilen puanların genellenebilirlik ve klasik test kuramına göre karşılaştırılması*. Yayımlanmamış doktora tezi. Hacettepe Üniversitesi, Ankara.
- Parsons, L. D. (2006). Using video to teach social skills to secondary students with autism. *Teaching Exceptional Children*, 39(2), 32-38.
- Pollaway, E. A., Serna, L., Patton, J. R., & Bailey, J. W. (2014). *Özel gereksinimi olan öğrenciler için öğretim stratejileri*. (Ş. Y. Özkan, Çev.). (10. Baskı). Ankara: Nobel Yayın.
- Rentz, J. O. (1987). Generalizability theory: a comprehensive method for assessing and improving the dependability of marketing measures. *Journal of Marketing Research*, 24(1), 19-28.
- Schoonen, R. (2005). Generalizability of writing scores: an application of structural equation modeling. *Language Testing*, 22(1), 1-30.
- Sharma, F., & Weathers, D. (2003). Assessing generalizability of scales used in cross-national research. *International Journal of Research in Marketing*, 20, 287-295.
- Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: a primer*. New-York: Springer.
- Sub, A., Valiga, M. J., & Goa, X. (1997). Using generalizability theory to assess the reliability of student ratings of academic advising. *Journal of Experimental Education*, 65(4), 367-379.
- Sucuođlu, B. ve Özokçu, O. (2005). Kaynaştırma öğrencilerinin sosyal becerilerinin değerlendirilmesi. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Özel Eğitim Dergisi*, 6(1), 41-57.
- Sudweeks, R. R., Reeve, S., & Bradshaw, W. S. (2005). A comparison of generalizability theory and Many-Facet Rasch Measurement in an analysis of college sophomore writing. *Assessing Writing*, 9(3), 239-261.
- Tse, J., Strulovitch, J., Tagalakis, V., Meng, L., & Fombonne, E. (2007). Social skills training for adolescents with Asperger syndrome and high-functioning autism. *Journal of Autism and Developmental Disorders*, 37(10), 1960-1968.

- Urbina, S. (2004). *Essentials of psychological testing*. New Jersey: John Wiley & Sons, Inc
- Volpe, R. J., McConaughy, S. H., & Hintze, M. H. (2009). Generalizability of classroom behavior problem and on-task scores from the direct observation form. *School Psychology Review*, 38(3), 382–401.
- Yelboğa, A. (2007). *Klasik test kuramı ve genellenebilirlik kuramına göre güvenilirliğin bir iş performansı ölçeği üzerinde incelenmesi*. Yayınlanmamış doktora tezi. Ankara Üniversitesi, Ankara.
- Yelboğa, A. (2008). Güvenirliğin değerlendirilmesinde genellenebilirlik kuramının kullanılması: endüstri ve örgüt psikolojisinde bir uygulama. *Psikoloji Çalışmaları Dergisi*, 28, 35-54.
- Yıldırım, İ. (2014). Bireyi tanıma teknikleri. Gürhan Can (Ed). *Psikolojik Danışma ve Rehberlik*. Ankara: Pegem Akademi.
- Yıldıztekin, B. (2014). *Klasik test teoremi ve genellenebilirlik kuramından puanlayıcılar arası tutarlılığın farklı yöntemlere göre karşılaştırılması*. Yayınlanmamış yüksek lisans tezi. Hacettepe Üniversitesi, Ankara.

## **EKLER DİZİNİ**



## EK 1. ETİK KURUL ONAY BİLDİRİMİ



T.C.  
HACETTEPE ÜNİVERSİTESİ  
Genel Sekreterlik

GIZLI

Sayı : 76000869/ 423-1525


14 Mayıs 2015

### EĞİTİM BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

İlgi: 28.04.2015 tarih ve 789 sayılı yazınız.

Enstitünüz Eğitim Bilimleri Anabilim Dalı Eğitimde Ölçme ve Değerlendirme Bilim Dalı tezli yüksek lisans programı öğrencisi **Zeynep PEKİN**'in öğretim üyesi **Yrd. Doç. Dr. Sevda ÇETİN**'in danışmanlığında yürüttüğü "**Otizm Sosyal Beceriler Profili Ölçeğinde Puanlayıcılar Arası Güvenirliğin Klasik Test Kuramı ve Genellenebilirlik Kuramındaki Farklı Desenlere Göre Karşılaştırılması**" konulu çalışma, Üniversitemiz Senatosu Etik Komisyonunun **12 Mayıs 2015** tarihinde yapmış olduğu toplantıda incelenmiş olup, etik açıdan uygun bulunmuştur.

Bilgilerinizi rica ederim.

  
Prof. Dr. Ömer UĞUR  
Rektör a.  
Rektör Yardımcısı

Ek: Tutanak

## EK 2. BİLGİ FORMU

Bu araştırma, otizmli çocukların ve gençlerin sosyal beceri düzeylerini belirlemek amacıyla gerçekleştirilmektedir. Lütfen aşağıdaki bilgi formunu ve ölçekleri dikkatle okuduktan sonra her soruyu yanıtlayınız. Yanıtlarınızda açık ve içten olmanız bu araştırma için çok büyük değer taşımaktadır. Vereceğiniz cevaplar yalnızca araştırma amacına yönelik olarak kullanılacak ve kesinlikle gizli tutulacaktır. Katıldığınız için teşekkür ederiz.

### ÇOCUK/GENÇ

Adı Soyadı :

Ölçeğin Doldurulduğu Tarih :

Doğum Tarihi :

Cinsiyeti :

Kardeş sayısı :

Tanısı :

Devam Ettiği Yaygın Eğitim Kurumu

(Anaokulu, İlkokul, Ortaokul, Lise)

Varsa Kurumun Adı :

Bu kuruma başlama tarihi :

Eğitime ilk başlama yaşı :

### ÖLÇEĞİ DOLDURAN

Adı Soyadı :



## EK 3. OTİZM SOSYAL BECERİLER PROFİLİ

(SCOTT BELLINI VE ANDREA HOPF)

### Yönerge:

Bu anket, otizimli bir çocuğun/gencin belli sosyal becerileri hangi sıklıkta sergilediğini ölçmek için hazırlanmıştır. Lütfen her soruyu dikkatlice okuyunuz ve bu çocuğun/gencin son bir ay ya da son iki ay boyunca sergilediği davranışları düşününüz. Çocuğun/gencin tanımlanan her davranışı hangi sıklıkta yaptığına karar veriniz.

Eğer çocuk/genç bir davranışı;

Hiçbir zaman yapmıyorsa: biri (1)

Bazen yapıyorsa: ikiyi (2)

Sık sık yapıyorsa: üçü (3)

Her zaman yapıyorsa: dördü (4) daire içine alınız.

Lütfen hiçbir maddeyi boş bırakmayınız. Bazı durumlarda çocuğun/gencin bir davranıştaki performansını gözlemleyememiş olabilirsiniz. Çocuğun/gencin bu davranıştaki olası performansını tahmin etmeye çalışınız.

<b>Sosyal Beceriler (sosyal karşılıklılık altboyutu)</b>	<b>Hiçbir zaman</b>	<b>Bazen</b>	<b>Sık sık</b>	<b>Her zaman</b>
1. Bir kişi hakkında sorular sorar.	1	2	3	4
2. Konular hakkında sorular sorar.	1	2	3	4
3. Konuşmalarda sıra alıp vermeyi sürdürür.	1	2	3	4
4. Başkalarına sempatisini ifade eder.	1	2	3	4
5. Başkalarının yüz ifadelerini tanır.	1	2	3	4
6. Başkalarının şakalarını veya esprilerini anlar.	1	2	3	4
7. Başkalarına yardım teklif eder.	1	2	3	4
8. Nasıl hissettiğini sözel olarak ifade eder.	1	2	3	4
9. Bir sohbeta bozmadan katılır.	1	2	3	4
10. Başkalarıyla selamlaşma başlatır.	1	2	3	4
11. Başkalarına iltifat eder.	1	2	3	4
12. Başkalarından nazikçe yoldan çekilmelerini ister.	1	2	3	4
13. Başkalarından gelen iltifatları kabul eder.	1	2	3	4
14. Kendisini bir başkasına tanıtır.	1	2	3	4
15. Uygun ses tonuyla konuşur.	1	2	3	4

## EK 4. GÖNÜLLÜ KATILIM FORMU

Bu çalışma, Dr. Sevda Çetin danışmanlığında Zeynep Pekin tarafından yürütülen bir tez çalışmasıdır. Araştırmanın amacı, otizmli çocukların sosyal beceri yetersizliklerini puanlayan katılımcıların puanlamadaki tutarlılıkları hakkında bilgi toplamaktır. Bu çalışma için Hacettepe Üniversitesi etik kurulundan izin alınmıştır. Çalışmaya katılım tamimiyle gönüllülük temelinde olmalıdır. Cevaplarınız tamimiyle gizli tutulacak ve sadece araştırmacılar tarafından değerlendirilecektir; elde edilecek bilgiler bilimsel yayımlarda kullanılacaktır.

Ölçek, genel olarak kişisel rahatsızlık verecek soruları içermemektedir. Ancak, katılım sırasında sorulardan ya da herhangi başka bir nedenden ötürü kendinizi rahatsız hissederseniz cevaplama işini yarıda bırakıp çıkmakta serbestsiniz. Böyle bir durumda çalışmayı yürüten kişiye, ölçeği tamamlamadığınızı söylemek yeterli olacaktır. Onay vermeden önce sormak isteyeceğiniz herhangi bir konu varsa sormaktan çekinmeyiniz. Bu çalışmaya katıldığınız için şimdiden teşekkür ederiz. Çalışma hakkında daha fazla bilgi almak için Hacettepe Üniversitesi Eğitimde Ölçme ve Değerlendirme Bilim Dalı öğretim üyelerinden Dr. Sevda Çetin (E-posta: [tsevda@hacettepe.edu.tr](mailto:tsevda@hacettepe.edu.tr)) ya da araştırma görevlisi Zeynep Pekin (E-posta: [zynppknn@gmail.com](mailto:zynppknn@gmail.com)) ile iletişim kurabilirsiniz.

***Bu çalışmaya tamamen gönüllü olarak katılıyorum ve istediğim zaman yarıda kesip çıkabileceğimi biliyorum. Verdiğim bilgilerin bilimsel amaçlı yayımlarda kullanılmasını kabul ediyorum.*** (Formu doldurup imzaladıktan sonra uygulayıcıya geri veriniz).

- Tarih:
- Katılımcı
  - Ad- Soyad:
  - Tel:
  - Adres:
  - İmza
- Araştırmacı:
  - Ad- Soyad:
  - Tel:
  - Adres:
  - E-posta:
  - İmza:

## EK 5. VELİ ONAY FORMU

Bu çalışma, Dr. Sevda Çetin danışmanlığında Zeynep Pekin tarafından yürütülen bir tez çalışmasıdır. Araştırmanın amacı, otizmli çocukların sosyal beceri yetersizliklerini doğal ortamlarında değerlendirilerek, puanlayıcıların puanlamadaki tutarlılıkları hakkında bilgi toplamaktır. Bu çalışma için Hacettepe Üniversitesi etik kurulundan izin alınmıştır. Katılım tamimiyle gönüllülük temelinde olmalıdır. Çocuklarınıza ilişkin bilgiler tamimiyle gizli tutulacak ve sadece araştırmacılar tarafından değerlendirilecektir; elde edilecek bilgiler bilimsel yayımlarda kullanılacaktır.

Ölçek, genel olarak çocuğunuza kişisel rahatsızlık verecek soruları içermemektedir. Onay vermeden önce sormak isteyeceğiniz herhangi bir konu varsa sormaktan çekinmeyiniz. Bu çalışmaya çocuğunuzun katılmasına izin verdiğiniz için şimdiden teşekkür ederiz. Çalışma hakkında daha fazla bilgi almak için Hacettepe Üniversitesi Eğitimde Ölçme ve Değerlendirme Bilim Dalı öğretim üyelerinden Dr. Sevda Çetin (E-posta: [tsevda@hacettepe.edu.tr](mailto:tsevda@hacettepe.edu.tr)) ya da araştırma görevlisi Zeynep Pekin (E-posta: [zynppknn@gmail.com](mailto:zynppknn@gmail.com)) ile iletişim kurabilirsiniz.

***Bu çalışmaya tamamen gönüllü olarak çocuğum katılmasına izin veriyorum ve istediğim zaman çocuğumun çalışmadan çıkabileceğini biliyorum. Verdiğim bilgilerin bilimsel amaçlı yayımlarda kullanılmasını kabul ediyorum.*** (Formu doldurup imzaladıktan sonra uygulayıcıya geri veriniz).

- Tarih:
- Veli:
  - Ad-Soyad:
  - Tel:
  - Adres:
  - İmza:
- Araştırmacı:
  - Ad- Soyad:
  - Tel:
  - Adres:
  - E-posta:
  - İmza:



Gösterge Paneli

Ödevler

Öğrenciler

Not Defteri

Kütüphaneler

Takvim

Tartışma

Teçhiller

GÖRÜNTÜLENİYOR: ANASAYFA > TEZ > ZEYNEP PEKİN TEZ

### Bu sayfa hakkında

Bu sizin ödev gelen kutunuzdur. Bir ödevi görüntülemek için, ödev başlığına tıklayın. Orjinallik Raporunu görmek için, benzelik kolonundaki orjinallik raporu ikonuna tıklayın. Bu ikon tıklanabilir durumda değilse, orjinallik raporu henüz oluşturulmamış demektir.

### zeynep pekin tez

GELEN KUTUSU | GÖRÜNTÜLENİYOR: YENİ ÖDEVLER ▼

Dosyayı Gönder

YAZAR

BAŞLIK

Ötzm Sosyal Beceriler Profili Ölçeğinde...

BENZELİK

%6

PUANLA

CEVAP

DOSYA

ÖDEV NUMARASI

55710634

TARİH

21-Haz-2015

GradelMark Raporu | Ödev ayarlarını düzenle

Telif Hakkı © 1988 - 2015 Paradigms, LLC. Tüm Hakları Saklıdır.

Kullanım Politikası | Geri Bildirim | Yardım Masası

## ÖZGEÇMİŞ

### Kişisel Bilgiler

<i>Adı Soyadı</i>	Zeynep Pekin
<i>Doğum Yeri</i>	Afyonkarahisar
<i>Doğum Tarihi</i>	27.01.1989

### Eğitim Durumu

<i>Lise</i>	Afyon Anadolu Öğretmen Lisesi	2007
<i>Lisans</i>	Yeditepe Üniversitesi Matematik Öğretmenliği	2013
<i>Yabancı Dil</i>	İngilizce: İyi Düzeyde	

### İletişim

<i>e-Posta Adresi</i>	zynppknn@gmail.com
<i>Jüri Tarihi</i>	29.05.2015