



Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü

Eğitim Bilimleri Anabilim Dalı

Eğitimde Ölçme ve Değerlendirme Bilim Dalı

**VOLEYBOL BECERİLERİ GÖZLEM FORMU İLE ELDE EDİLEN  
PUANLARIN GENELLENEBİLİRLİK VE KLASİK TEST  
KURAMINA GÖRE KARŞILAŞTIRILMASI**

Meltem Emel Öztürk

Yüksek Lisans Tezi

Ankara, 2011

VOLEYBOL BECERİLERİ GÖZLEM FORMU İLE ELDE EDİLEN  
PUANLARIN GENELLENEBİLİRLİK VE KLASİK TEST KURAMINA  
GÖRE KARŞILAŞTIRILMASI

Meltem Emel Öztürk

Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü

Eğitim Bilimleri Anabilim Dalı

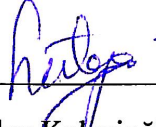
Eğitimde Ölçme ve Değerlendirme Bilim Dalı

Yüksek Lisans Tezi

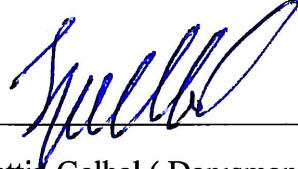
Ankara, 2011

**KABUL VE ONAY**

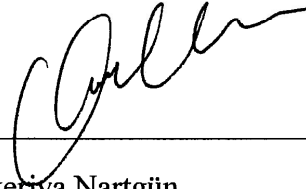
Meltem Emel Öztürk tarafından hazırlanan “Voleybol Becerileri Gözlem Formu ile Elde Edilen Puanların Genellenebilirlik ve Klasik Test Kuramına Göre Karşılaştırılması” başlıklı bu çalışma, 08.07.2011 tarihinde yapılan savunma sınavı sonucunda başarılı bulunarak jürimiz tarafından Yüksek Lisans Tezi olarak kabul edilmiştir.



Doç. Dr. Hülya Kelecioğlu ( Başkan )



Prof. Dr. Selahattin Gelbal ( Danışman )



Doç. Dr. Zekeriya Nartgün

Yukarıdaki imzaların adı geçen öğretim üyelerine ait olduğunu onaylarım.

Prof. Dr. İrfan Çakın

Enstitü Müdürü

## BİLDİRİM

Hazırladığım tezin/raporun tamamen kendi çalışmam olduğunu ve her alıntıya kaynak gösterdiğimi taahhüt eder, tezimin/raporumun kağıt ve elektronik kopyalarının Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü arşivlerinde aşağıda belirttiğim koşullarda saklanmasına izin verdiğimi onaylarım:

- Tezimin/Raporumun tamamı her yerden erişime açılabilir.
- Tezim/Raporum sadece Hacettepe Üniversitesi yerleşkelerinden erişime açılabilir.
- Tezimin/Raporumun ...3... yıl süreyle erişime açılmasını istemiyorum. Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin/raporumun tamamı her yerden erişime açılabilir.

08.07.2011



---

Meltem Emel Öztürk

CANIM ABLAM N. AYŞE ÖZTÜRK'E

## TEŞEKKÜR

Öğrenimime başladığım günden bu yana beni her konuda destekleyen, her türlü imkanı sunarak mükemmel bir çalışma ortamı sağlayan, öneri ve yorumlarıyla bakış açımı genişleterek bu çalışmanın tamamlanmasında büyük katkıları olan saygıdeğer hocam ve danışmanım Prof. Dr. Selahattin GELBAL'a teşekkürlerimi sunarım.

Akademik anlamda yetişmemde büyük emeği olan, deneyimleriyle ve düşünceleriyle beni her anlamda zenginleştiren, çalışmalarımda bana yol gösterici olan, ilgisini ve desteğini esirgemeyen saygıdeğer hocam Doç. Dr. Hülya KELECİOĞLU'na, tez konumu belirlemeye çalıştığım ilk aşamadan itibaren fikrin kafamda canlanıp gelişmesini sağlayarak her türlü heyecanımı paylaşan, yardımları ve önerileriyle çalışmama katkıda bulunan saygıdeğer hocam Doç. Dr. Nuri DOĞAN'a teşekkür ederim.

Bu çalışmada bana yol gösterici olan ve desteğini esirgemeyen, verilerin analiz aşamasında her türlü yardımı sağlayan, güler yüzü ve sınımsız gülüşüyle bana pozitif enerji aşılayarak tezimin ortaya çıkmasını sağlayan Sayın Yrd. Doç. Dr. Neşe GÜLER'e şükranlarımı sunarım.

Yaşamımın her devresinde olduğu gibi yüksek lisansa başladığım andan itibaren de her zaman yanımda olup, bütün sorunlarımın üstesinden gelmeme yardımcı olan, belki de beni benden daha fazla düşünen sevgili ablam N. Ayşe ÖZTÜRK'e, benden uzakta olsa da her zaman yanımda olduğunu hissettirerek bana her konuda destek olmaya çalışan sevgili ablam Prof. Dr. S. Neşe ATABEY'e, en zor zamanlarımda sen her şeyi başarırısın diyerek beni motive eden ve her türlü yardımı yapan sevgili ablam Uz. Dr. Özden Ş. ÜNERİ'ye teşekkür ederim.

Hayatımın her anında beni destek ve sevgilerinden yoksun bırakmayarak her zaman ileriye gidebilmem için beni yüreklendiren ve bu günlere gelmemi sağlayan canım anne ve babama çok teşekkür ederim.

Burada adlarını sayamadığım, tez yazım sürecinde çeşitli şekillerde desteklerini gördüğüm çalışmam sırasında beni yalnız bırakmayan sevgili arkadaşlarıma ve emeği geçen herkese teşekkürlerimle...

## ÖZET

ÖZTÜRK, Meltem Emel. *Voleybol Becerileri Gözlem Formu ile Elde Edilen Puanların Genellenebilirlik ve Klasik Test Kuramına Göre Karşılaştırılması*, Yüksek Lisans Tezi, Ankara, 2011.

Bu araştırmada, voleybolcuların teknik becerilerinin tespit edilmesi için kullanılan ‘Voleybol Becerilerine İlişkin Gözlem Formu’ (VBGF) ndan elde edilen ölçmelerin; Klasik Test Kuramı (KTT) ve Genellenebilirlik (G) Kuramına dayalı olarak belirlenen güvenilirlik katsayılarının düzeyleri saptanmış ve birbirleriyle karşılaştırılmıştır.

Araştırmanın çalışma grubunu, 2009-2010 yılları arasında Türkiye Voleybol Federasyonu tarafından belirlenmiş olan 10 bölgede (Ankara- İstanbul- İzmir- Bursa- Çanakkale- Konya- Eskişehir- Antalya- Trabzon- Diyarbakır) görevlendirilen antrenörlerin seçmiş olduğu 102 erkek voleybolcu oluşturmuştur. Voleybolcuların teknik becerilerinin (pas, manşet, smaç, blok ve voleybol zekası) ölçülmesi amacıyla 5 görevden oluşan “Voleybol Becerilerine İlişkin Gözlem Formu” kullanılmıştır. Gözlem Formuna ait maddeler, beş dereceli ölçekte puanlayıcılardan (4 puanlayıcı), “Görülmedi”, “Yetersiz”, “Yeterli”, “Oldukça Yeterli” ve “Mükemmel” seçeneklerinden birini işaretlemeleri istenmiştir.

VBGF’nun toplamından elde edilen veriler Klasik Test Kuramında puanlayıcıların iç tutarlılık güvenilirliği için Cronbach  $\alpha$  (alfa) katsayısı, puanlayıcılar arasındaki tutarlılık için Kendall uyum katsayıları hesaplanmıştır. Genellenebilirlik Kuramında, b (birey), g (görev) ve p (puanlayıcı) olmak üzere bxgxp tümüyle çaprazlanmış desen kullanılarak genellenebilirlik ve güvenilirlik katsayıları hesaplanmıştır.

VBGF’nun toplamından elde edilen puanların güvenilirliği için puanlayıcı sayılarının 1, 2, 3, 4 ve 5 ve görev sayılarının 2, 5, 8 ve 10 olduğu koşullarda, G kuramında ana ve

ortak etkilerin varyans bileşenlerin kestirilmesi için SPSS 16.0 programında G ve Phi kat sayıları bulunmuştur. Alt boyutlara ait sonuçlar incelendiğinde bütün alt boyutlar için en yüksek varyans bileşenin gxp ortak etkisi için kestirilen varyans bileşeni olduğu görülmektedir. Bunu bxgxp ortak etkisi varyans bileşeni takip etmektedir. Bunların dışındaki temel ve ortak etkilere ait varyans bileşenlerinin ise sifıra yakın olduğu görülmektedir. Bu sonuca göre, tek değişkenli analizle bu ölçme için görev x puanlayıcı ortak etkisinden kaynaklanan farklılığın çok büyük olduğu, belli görevlere verilen puanların puanlayıcıdan puanlayıcıya çok fazla değiştiği görülmektedir.

VBGF puanlamasından elde edilen güvenilirlik katsayıları iç ölçütlere göre incelendiğinde G katsayısı ve Cronbach Alfa katsayıları beklenen değerlerinden düşük olduğu görülmektedir. Ayrıca, Phi katsayısı ile Kendall uyum katsayılarının da beklenen değerlerinden düşük olduğu görülmektedir.

Elde edilen sonuçlara göre voleybol becerilerine ilişkin dereceleme ölçeğinde aynı ölçme durumu için Klasik Test Kuramında ve Genellenebilirlik Kuramının çok değişkenli modelle elde edilen güvenilirlik katsayılarının birbirleriye uyumlu sonuçlar ürettiği görülmüştür. Bununla birlikte, VBGF'nun güvenilirliğinin düşük olması, ölçeğin güvenilir bir ölçek olarak değerlendirilemeyeceğini göstermektedir.

### **Anahtar Sözcükler**

Klasik test kuramı, Genellenebilirlik kuramı, Performans ölçümü, Güvenirlik.

## ABSTRACT

ÖZTÜRK, Meltem Emel. *The Comparison of reliability of the “Volleyball Abilities Observation Form” (VAOF) points to the Generalizability and the Classical Test Theory*, Master Thesis, Ankara, 2011.

In this study, the measurements which were obtained from the observation form related to the volleyball abilities, were used in order to determine the technical abilities of volleyball players; Classical Test theory and the levels of reliability coefficients stated at the base of the Generalizability Theory were fixed and these levels were compared among them.

The Research group consists of the 102 boy volleyball players who were chosen by the trainers given the task by Turkish Volleyball Federation in 10 region (Ankara- İstanbul- İzmir- Bursa- Çanakkale- Konya- Eskişehir- Antalya- Trabzon- Diyarbakır) between 2009-2010. By the aim of the measuring technical abilities of the volleyball players (set, spike, hitting, blocking and intelligence of volleyball), the observation form which was composed of 5 job items, was used. It was requested that the four examiners should mark one of the options; “not seen”, “deficient”, “sufficient”, “Considerably sufficient” and “Perfect” for the items in the *VAOF*, using a five graded scale

The data which were obtained from *VAOF* were calculated in the classical test theory for the internal consistency reliability of the elements, Cronbach  $\alpha$  (alfa) coefficient, and for the consistency among the raters, Kendall concordance coefficients were figured out. In generalizability theory, by using  $p \times t \times r$  (all facet cross with each other) design, generalizability and dependability coefficient were calculated.

For the reliability of the grades that were obtained from the entire and subdimensions of the *VAOF* related to the volleyball abilities, in the conditions of the slightly decrease and increase of the element numbers and existency of the rates' numbers as 1,2,3,4,5 and task numbers as 2,5,8,10 in the G theory in order to estimate the variance components of main and common effects, the G and Phi coefficients are found. When

the results belonging the sub-dimensions are analysed, it is seen that for the all sub-dimensions, the highest variance component is the variance component which is estimated for the individual main influence. This is followed by bxgxp common influence variance component. Except these, it is seen that the other variance components belonging to the main and common influences are closed to zero.

When the reliability coefficients that are obtained from the grading of Gradation Scale concerning the volleyball ability are examined upon the internal criteria, it was seen that G co-efficient and Cronbach alfa reliability coefficients are lower than the expected values. At the same time, Phi and Kendall reliability coefficients are lower than the expected values.

According to the obtained results, it is seen that, for the same measurement condition for the *VAOF*, the reliability coefficients obtained from the Classical Test Theory and from the Multivariate Model of Generalizability Theory produced harmonious results. Nonetheless, being low of reliability of the Volleyball ability rating coefficient's joint observation form and sub-dimensions of the observation form indicated that the *VAOF* could not be evaluated as a confidential form.

### **Key Words**

Classical Test Theory, Generalizability Theory.Measurement of Performance, Realibility

## İÇİNDEKİLER

	Sayfa No
KABUL VE ONAY.....	i
BİLDİRİM.....	ii
ADAMA.....	iii
TEŞEKKÜR.....	iv
ÖZET.....	v
ABSTRACT.....	vii
İÇİNDEKİLER.....	ix
TABLolar DİZİNİ.....	xii
ŞEKİLLER DİZİNİ.....	xiii
BÖLÜM I.....	1
GİRİŞ.....	1
1.1. PSİKOMOTOR BECERİLERİ VE PUANLAMA ARAÇLARI.....	11
1.1.1. İşaretleme Tipi Araçlar.....	11
1.1.1.1. Kontrol Listeleri.....	11
1.1.1.2. Derecelendirme Ölçekleri.....	12
1.1.1.2.1 Sayısal Derecelendirme Ölçeği.....	13
1.1.1.2.2. Betimsel Derecelendirme Ölçeği.....	14
1.1.1.2.3. Karşılaştırmalı Derecelendirme Ölçeği.....	14
1.1.1.2.4. Grafik Derecelendirme Ölçeği.....	15
1.1.2. Rubrikler.....	15
1.1.2.1. Holistik Rubrikler.....	15
1.1.2.2 Analitik Rubrikler.....	16
1.2. PERFORMANS DEĞERLENDİRMEDE GÜVENİRLİK.....	17
1.3. KLASİK TEST KURAMINA GÖRE GÜVENİRLİK.....	18
1.4. GENELLENEBİLİRLİK KURAMINA GÖRE GÜVENİRLİK.....	25
1.4.1. Genellenebilirlik (G) Çalışmalar (G–Study).....	29

1.4.2. Karar (D) Çalışmaları (D–Study) .....	30
1.4.3. Bağlı Hata ve Mutlak Hata .....	31
1.4.4. Genellenebilirlik Katsayısı (G–Katsayısı) ve Güvenirlilik İndeksi (Phi–katsayısı) .....	32
1.4.5. Tesadüfi Değişkenlik Kaynağı ve Sabit Değişkenlik Kaynağı .....	32
1.4.6. Çaprazlanmış Desen ve Yuvalanmış Desen .....	33
1.4.7. Değişkenlik Kaynağı Olarak Puanlayıcılar .....	34
1.5. PROBLEM DURUMU .....	38
1.6. PROBLEM CÜMLESI .....	39
1.7. ALT PROBLEMLER .....	40
1.8. ARASTIRMACININ AMACI .....	40
1.9. ARASTIRMANIN ÖNEMİ .....	41
1.10. SAYILTILAR .....	41
1.11. SINIRLILIKLAR .....	41
1.12. KISALTMALAR .....	42
1.13. İLGİLİ ARASTIRMALAR .....	42
1.13.1. Yurt İçinde Yapılan Araştırmalar .....	42
1.13.2. Yurt Dışında Yapılan Araştırmalar .....	46
BÖLÜM II .....	51
YÖNTEM .....	51
2.1. ARASTIRMANIN TÜRÜ .....	51
2.2. ÇALIŞMA GRUBU .....	51
2.3. VERİ TOPLAMA ARACI .....	51
2.4. VERİLERİN TOPLANMASI .....	53
2.5. VERİLERİN ANALİZİ .....	54
BÖLÜM III .....	57
BULGULAR VE YORUM .....	57
3.1. KLASİK TEST KURUMUNA GÖRE VOLEYBOL BECERİLERİ PERFORMANSININ DEĞERLENDİRİLMESİNİN İNCELENMESİ .....	57

3.1.1. Klasik Test Kuramına Göre Voleybol Becerileri Performansının Değerlendirilmesinin İç Tutarlılığı .....	57
3.1.2. Klasik Test Kuramına Göre Voleybol Becerileri Performansının Değerlendirilmesinde Dört Farklı Puanlayıcının Puanları Arasındaki Tutarlılık Düzeyi .....	58
3.2. GENELLENEBİLİRLİK KURAMINA GÖRE VOLEYBOL BECERİLERİ PERFORMANSININ DEĞERLENDİRİLMESİNİN İNCELENMESİ .....	60
3.2.1. Voleybol Becerileri Performansının Değerlendirilmesi Kestirilen Genellenebilirlik Kuramı Parametreleri .....	60
3.2.2. Orjinal Puanlayıcı ve Görev Sayıları İle Puanlayıcı Sayılarının Arttırılıp Azaltılarak 1, 2, 3, 4, 5 Oldukları ve Görev Sayılarının Arttırılıp 2, 5, 8, 10 Oldukları Senaryoları Sonucunda K Çalışmasıyla Kestirilen, G ve Phi Katsayıları .....	63
BÖLÜM IV .....	67
SONUÇ VE ÖNERİLER .....	67
4.1. SONUÇLAR .....	67
4.2. ÖNERİLER .....	69
KAYNAKÇA .....	72
EKLER .....	79
ÖZGEÇMİŞ .....	81

**TABLULAR DIZINI****Sayfa No**

<b>Tablo 1.</b> İki Değişkenlik Kaynaklı Tesadüfi Desen İçin Varyans Bileşenlerinin Kestirilmesi.....	35
<b>Tablo 2.</b> Sporcu Puanlarının Dört Puanlayıcıya İlişkin Betimsel İstatistikleri .....	57
<b>Tablo 3.</b> Sporcu Puanlarının Dört Puanlayıcıya İlişkin İç Tutarlılığı.....	58
<b>Tablo 4.</b> Dört Puanlayıcının Beş Göreve Verdikleri Puanlar Arasındaki Korelasyon Katsayıları.....	59
<b>Tablo 5.</b> Voleybol Becerileri Performansının Değerlendirilmesinin G Çalışması Sonucunda Değerlendirmenin Kestirilen Varyansları ve Toplam Varyansı Açıklama Oranları.....	61
<b>Tablo 6.</b> Voleybol Becerileri Performansının Değerlendirilmesinin K Çalışması İle Madde ve Puanlayıcı Sayıları Senaryolarına Göre G ve Phi Katsayıları.....	63

## SEKİLLER DİZİNİ

### Sayfa No

<b>Sekil 1.</b> Voleybol Becerileri Performansının Değerlendirilmesinin K Çalışması İle Madde ve Puanlayıcı Sayıları Senaryolarına Göre G Katsayısı.....	65
<b>Sekil 2.</b> Voleybol Becerileri Performansının Değerlendirilmesinin K Çalışması İle Madde ve Puanlayıcı Sayıları Senaryolarına Göre Phi Katsayısı.....	65

# BÖLÜM I

## GİRİŞ

Bilimsel çalışmaların yapılmasında vazgeçilmez bir yeri olan ölçme, kuramlarda yer alan teorik ilişkilerin, gerçek hayatta test edilmesini sağlayarak, bilimsel çalışmaların tartışmasız en önemli parçası olmaktadır. Ölçme aynı zamanda, bilimin işlevsel sürecinde yer alan, aynı amaca yönelik oluşturulan kuramların aynı şartlar altında kıyaslanarak, bilgi sağlama açısından çalışan ve çalışmayan yanlarının saptanması, daha pratik ve doğru olanının belirlenmesi, kuramların benzer ve farklı yönlerinin açığa çıkarılmasında önemli bir yere sahiptir (Güler,2008).

Ölçmenin, çalışmaların yapıldığı alanlara göre, literatürde birçok tanımı yer almaktadır. Genel olarak tanımlamak gerekirse ölçme; nesnelere ya da bireylere, belirli bir özelliğe sahip oluş derecelerini belirtmek için, belirli kurallara uyarak sembolik değerler verme işlemidir (Turgut, 1993). Ölçme sonuçları tek başına anlamlı değildir. Bu sonuçları anlamlı hale getirebilmek için değerlendirmeye ihtiyaç vardır. Değerlendirme ise, ölçme sonuçlarını bir ölçüt veya ölçütler takımı ile karşılaştırarak birey ya da nesnelere ölçülen özellikleri hakkında bir karar verme sürecidir (Baykul, 2000; Özçelik, 1998). Kullanılan ölçütün uygunluğunun yanı sıra, ölçme sonuçlarının güvenilir ve geçerli olmasının değerlendirmenin doğru yapılmış olması açısından önemi büyüktür. Değerlendirme sonuçlarının isabetli olma derecesini arttırabilmek için, yapılan ölçme işlemlerinde kullanılan ölçme araçlarının güvenilirliğinin ve geçerliğinin olabildiğince yüksek olması istenmektedir (Alharby, 2006). Bir bilim dalındaki çalışmalar ve bunların uygulamaya konulması, o bilim dalına özgü ölçme araç ve yöntemlerinin bulunmasıyla hızlanmıştır. Ölçme bilim için o kadar önemli hale gelmiştir ki, çeşitli bilim dalları için ölçme teorileri geliştirilmiştir (Baykul, 2000).

Spor bireyin tabii çevresini beşeri çevre haline çevirirken elde ettiği yetenekleri geliştiren, belirli kurallar altında araçlı ya da araçsız, bireysel veya takım halinde faaliyet kapsamı içinde veya tam zamanını alacak şekilde meslekleştirilerek yaptığı sosyalleştirici, toplumda bütünleştirici, ruh ve fiziği geliştiren rekabetçi ve dayanışmacı kültürel bir olgudur (Erkal,1981). Bugünün dünyasında, ülkeler hangi spor dalında olursa olsun uluslar arası sportif karşılaşmalarda ön sıralarda olmayı ve şeref kürsüsüne

çıkmayı amaç edinmişlerdir. Kamuoyunun gençler arasındaki yeteneğin erkenden belirlenmesi konusunda büyük beklentileri vardır. Sporda yeteneğin, çok erken yaşlarda, düşük değerli kriterlerle seçiminin riskli olduğu, bunun ise yeteneği tam belirlenmezse belki de bireyin spordan uzaklaşmasına ve spora yetenekli bir bireyin kaybolmasına yol açacağı düşünülmektedir. Araştırmalar; sporda en iyi motivasyonun gençler için spor aktivitelerinin iyi şekilde organizasyonu ile gerçekleştirileceğini ortaya koymaktadır.

Birçok alanda olduğu gibi sporun da hem kuramsal hem de deneysel yönleri bulunmaktadır. Sporda, beceri, yetenek, performans gibi yapıların incelenmesinde, içeriklerinin açıklanmasında ve bireylerin bu niteliklerinin sayılarla ifade edilmesinde ölçmeden yararlanılır. Ölçme ve ölçme araçları sporun vazgeçilmez bir parçası olan performans değerlendirmenin temelini oluşturmaktadırlar. Ölçme gerek günlük yaşantımızda, gerekse bilimsel çalışmalarda önemli bir yer tutar. Birçok bilim dalındaki ilerlemeler ölçme araç ve yöntemlerinin geliştirilmesiyle sağlanabilmiştir. Ayrıca, ölçme araç ve yöntemleri bilimsel bulguların uygulama alanına konulabilmesine önemli ölçüde katkı getirmiştir. Bugün ölçmeyi kullanmayan bir bilim dalı ve duyarlı ölçme yapan araçlara dayanmayan bir teknoloji yok gibidir. Ölçmenin diğer gözlem tekniklerine üstünlüğü daha kesin, daha nesnel ve daha güvenilir bir yol olmasıdır. Toplanan gözlem verilerinin sayı cinsinden ifade edilmiş olması istatistiksel çözümleme ve yorumlamayı kolaylaştırır, hükümde daha kesin ve açık olmayı sağlar.

Duyarlı ölçmeler yapmada araç kullanma ihtiyacı vardır; bu ihtiyaç psikoloji ve eğitim gibi, değişkenleri çoğu zaman doğrudan gözlenemeyen bilim dallarında daha da fazladır. Bilimin konusu, kuramsal yapı ve elemanlarla bunlara ilişkin verileri olan bir sistemin elemanları arasındaki bağlantıları ortaya koymak, bu bağlantıları genellemek ve bunlardan yola çıkarak kanunlara varmaktır. Kuramsal çalışmalarda, kuramsal yapıdan elde edilen ilişkilerin pratikte doğruluğunu yoklamada; deneysel çalışmalarda, nitelikleri niceleyip, bunlardan yola çıkarak ilke ve genellemelere ulaşmada, bu genellemelerin değişik durumlarda doğru olması halinde kanunlara varmada ölçmeden yararlanılır. Ayrıca, bireylerin ve grupların belli özellikler yönünden karşılaştırılmasında da ölçmeye başvurulur.

Sporda beceri, her hareketin birbirini doğru olarak izleyen şekilde ve istenilen kuvvetle meydana gelmesi olayıdır. Becerili (koordineli) hareket, merkezi sinir sisteminden, kasılması gereken kaslara gerektiği zamanda uyarıların gelmesiyle olur. Böylece beceri ve teknik ile bütünlenen performans için gereken hareketler yapılır (Açıkada, Ergen, 1990). Voleybolda beceri; başarıyı etkileyen en önemli faktörlerden biridir. Bir diğer ifadeyle; beceri, motorik özelliklerin bileşkesi olan kondisyon ve taktikle birlikte performansa etki eden faktörlerden biridir (Tiryaki, 1999).

Performans, fiziksel çabalara uyabilme ve onlara uygun cevap verebilme kapasitesi olarak tanımlanabilir. Performans ve uygunluk kişinin enerji kullanma kapasitesi (aerobik ve anaerobik şartlarda), nöromüsküler fonksiyonu (kas kuvveti, endürans ve teknik) ve psikolojik faktörlerle (kişinin motivasyonu ve taktikler ) ölçülerek değerlendirilir. Ayrıca performans kişinin uğraştığı spor dalında ulaştığı en yüksek başarı seviyesidir (Ergun ve Baltacı, 1997).

Baysal (1999)'a göre performans, bir kişinin sahip olduğu potansiyel veya gerçek bilgi, yeteneklerini amaçlarına veya beklentilerine ulaşabilmek için ne ölçüde kullanabildiğini tanımlayan bir kavramdır. Bir başka deyişle, kişinin sahip olduğu kapasitenin, bir işi belli zaman dilimi içinde başarıyla tamamlamada kullanabilme yüzdesidir. Performans, bir fiziksel aktivite sırasında, o fiziksel aktivitenin gerektirdiği fizyolojik, biyomekanik ve psikolojik verim olarak da tanımlanmaktadır (Kuter ve Öztürk,1997).

Performans ölçmenin sporcu başarısı için öneminin artması, bu kavram ile ilgili çalışmaların artmasına neden olmaktadır. Bu araştırmalardan elde edilen genel yöntemlere bakıldığında; Performansın (Psikomotor beceriler) ölçülmesinde şu genel yöntemden birine başvurulabilir;

- Birey, hedef alınan beceriyi yapar; puanlayıcı, uygulama sırasında bireyin hareketlerini gözler ve yapılan uygulamayı puanlar. Bu yöntem, aslında, davranışın gözlenip ölçülmesidir.
- Birey, hedef alınan beceriyi yaparak bir beceri ürünü ortaya çıkarır; puanlayıcı, bu ürünü inceler ve ona puan verir. Bu yöntemde ölçülen şey,

ürünün mükemmelliğidir. Aslında, ürüne bakarak bireyin becerisi hakkında bir değer yargısına varılır.

- Birey, hedef alınan beceriyi yaparak bir ürün meydana getirir. Puanlayıcı, hem bireyin davranışlarını hem de ürünün oluşmasını, çeşitli basamaklarda gözler ve her basamakta ara puanları takdir eder, sonunda da bir toplam puana varır.

Bu yöntemlerden hangisinin kullanılacağı, hedefin niteliğine ve ölçmenin amacına bağlıdır. Hedef, öğrencinin bir milli oyunu oynayabilmesi ise, ortaya çıkacak maddi bir ürün bulunmadığından, birinci yöntem kullanılır. Hedef, bir öğrencinin ağaç işlerindeki becerisi ise, ondan bir masa yapması istenebilir. Bu halde, masaya verilecek puanla yetinilecekse, ikinci yöntem kullanılır. Masaya verilecek puan, hedef davranışları yeterince temsil etmiyorsa, işin ara basamaklarına ve öğrencinin işi yaparken gösterdiği davranışlara da puan vermek gerekeceğinden, üçüncü yöntem kullanılır (Turgut, F. 1977).

Bireyler hakkında farklı yollarla bilgi toplamak mümkündür. Bu yollar test teknikleri ve test dışı teknikler olarak gruplandırılabilir. Tekin (2000) test tekniklerini; “tipik davranış testleri” ve “maksimum performans testleri” olarak ikiye ayırmıştır.

Maksimum performans testlerinde, bir kişinin bir işi belli bir durumda ne denli iyi ya da doğru yapabildiği belirlenmeye çalışılır. Birey hem kendisi hem de başkaları ile yarış durumundadır; yapabileceğinin en fazlasını, en iyisini yapmaya çalışır. Bu testler; zeka testleri, başarı testleri, psikomotor davranışların ölçüldüğü testler, müzik yeteneği testi gibi yetenek ve başarı testlerini içerir (Erkuş, 2000; Tekin, 2000).

Tipik davranışın ölçülmesinde, kişinin belli bir durumdaki davranışlarının ve bu durumdaki özelliklerinin belirlenmesi ile ilgilenilir. Burada doğru ya da yanlış cevap değil, doğru bir şekilde davranışla ilgili olduğu düşünülen özelliklerin belirlenmesi önemlidir. Tutum, ilgi ve kişilik gibi değişkenlerin ölçülmesi bu gruba girmektedir (Erkuş, 2000; Tekin, 2000).

Psikomotor beceriler belirli fiziksel hareketlerin belli bir sıraya göre doğru, hızlı ve otomatik olarak yapılması sonucunda ortaya çıkan davranışları içerir. Bir müzik aletini çalma, tenis oynama, yemek pişirme, daktiloda on parmak yazı yazma, resim çizme, bandaj yapma, bir video kamerasını kurma ve çalıştırma gibi davranışlar psikomotor becerilere örnek olarak verilebilir. Performansın (Psikomotor beceriler) önemli olduğu voleybol, dans, jimnastik vb. gibi spor branşları geniş ölçüde vücut hareketlerine dayanır. Psikomotor becerileri ölçmenin oldukça güç olduğu ve bu tür ölçmede kullanılacak araç ve yöntemlerin geliştirilmesinde yetersizliklerin bulunduğu görülmektedir (Turgut,1993). Psikomotor beceriler olarak tanımlanan spor dallarında performansı ölçme; sporcunun fiziksel, fizyolojik, biyomotorik ve psikososyal özelliklerinin bileşkesinin sonucu olarak ortaya koyduğu verim düzeyi şeklinde açıklanmaktadır (Açıkada ve Ergen, 1990). Spor branşlarında, başarılı olan sporculardaki yapısal özellikleri tespit edilerek, fiziksel yetenek ve spor dalı arasındaki ilişkilere göre belirlenen ve yaşa bağlı olarak değişmeyen ideal sporcucu modellerinin araştırılması özellikle son yıllarda gündemini korumaktadır (Gürses ve Olgun, 1989).

Psikomotor becerileri ölçme, hedeflerle ilişkili olarak yapılmalıdır. Ölçme, hedefi davranışlarıyla tanımlayacak kadar ayrıntılara indirilmelidir. Böylece listelenen davranışlar gözlenebilir ve puanlanabilir nitelikte olacaktır. Hedefe ne derecede ulaşıldığını yoklamak için o hedefi oluşturan davranışların tümünün ölçülmesi gerekmez. Beceriye oluşturan davranışlar zincirinde, ancak kritik davranış halkaları gözlenip ölçülmelidir (Darst ve ark., 1989, akt. Deliceoğlu, 2009).

Psikomotor beceriler, karmaşık davranışlar zincirinden oluşur. Bir becerinin ölçülmesi için, bu zincirin önemli halkaları gözlenebilmeli ve gözlem sonuçları sayısallaştırılmalıdır. Psikomotor beceriyi oluşturan davranışları gözleyerek ölçmede izlenecek yöntemin ana basamakları aşağıdaki şekilde olmalıdır:

1. Beceriye oluşturan davranışlar iyice analiz edilmeli, bunlar arasından ölçme için gerekli kritik davranışlar saptanmalıdır.
2. Kritik davranışların gözlenmesini ve puanlanmasını kolaylaştıracak bir araç hazırlanmalıdır.

3. Bireyin, saptanan kritik davranışları yapabileceği gözlem ve sınaama durumları yaratılmalıdır.
4. Birey, oluşturulan bu durumlarda davranışı yaparken gözlenmeli gözlenen davranışlar puanlanmalıdır (Turgut, 1993).

Kritik davranışların gözlenip puanlanması için bir araç hazırlama işi, hedefe ait davranışların analizi iyi yapılırsa kolaylaşır. Analiz sonucunda seçilen kritik davranışlar, kullanılışlı bir düzenle araçta yer almalıdır. Bir becerinin uygulanmasındaki basamaklara göre sıralama, çoğu halde kullanılışlı bir araç meydana getirir. Hazırlanan ölçme aracında puanlamanın nasıl yapılacağı da açıkça gösterilmelidir. Bireyin kritik davranışları uygulayacağı durum ya gerçek bir durum olarak seçilir ya da test için özellikle hazırlanır. Örnek olarak; bireylerin, voleybol oynama becerileri gerçek bir maçta gözlenebileceği gibi test için düzenlenmiş bir oyunda da gözlenebilir (Turgut, 1993:163).

Psikomotor becerileri ölçme işlemlerinde, beceriler görüntü olarak kaydedilir ve yavaşlatılarak davranışın ayrıntıları iyice gözlenir. Bir kez uygulanan hareketlerde, özellikle hareket hızlı ve karmaşık ise, bütün davranışları birlikte gözlemenin zorluğunu engellemek amacıyla beceri video kameradan tekrarlatılarak her seferinde belli bir kritik davranış gözlenir (Welk, 2002).

Hatasız bir ölçüm elde etmek ölçme yapan kişinin ilk ve en önemli amacıdır. Ancak hatasız bir ölçüm elde etmek, çok kolay değildir. Elle tutulur, gözle görülür ve kararlı özellikler üzerindeki bir başka deyişle somut ölçümlerin bile tümüyle hatadan arınmış olması olanaksızdır. Aynı koşullarda fakat birbirinden bağımsız olarak tekrarlanan ölçme sonuçlarından hesaplanan değer ile ölçülen özelliğin gerçek değeri arasındaki fark ölçme hatası olarak adlandırılır (Özçelik,1998).

Ölçmenin standart hatası, bir bireyin sonsuz sayıda tekrarlı ölçmelerinde gözlenen ve gerçek puanı arasındaki uyumsuzluğun standart sapmasıdır (Crocker ve Algina, 1986). Ölçmelere karışan çeşitli hatalar nedeniyle ölçülen özelliğin kestirilen gerçek değeri bulunamaz. Üzerinde ölçme yapılan değişkenin gerçek puanı, ölçme aracı, ölçmeyi

yapan kişi ve ölçme ortamı gibi değişkenlerin sebep olduğu hatalardan etkilenmektedir. Ölçmelerde hata, ölçülmek istenmeyen miktarın ölçme sonuçlarına karışmasıdır. Ölçme ve ölçme sonuçlarına, sabit, sistematik ve rastlantısal (tesadüfi) hata olarak üç türde hata karıştığı görülmektedir. Sabit hatalar, ölçme sonuçlarına aynı yönde ve miktarda karışan hatalardır. Bu tür hatalar ölçek iyi kalibre edilmediğinde oluşur. Sistematik hatalar, belirli bir yönde ilerleyici ya da gerileyici miktarlarda ortaya çıkan hatalardır. Ölçülen büyüklüğe, ölçmeciyeye ya da ölçme koşullarına bağlı olarak miktarı değişen hatalardır. Bu hatalara yanlılık adı da verilir. Kontrol edilebilirdir. Rastlantısal (tesadüfi) hatalar, değişen kişisel koşullar, çevre koşulları ya da ölçüm aracındaki değişiklikler nedeniyle şansa ortaya çıkan ve hangi yönde etki ettiği kestirilemeyen hatalardır. Klasik test kuramı bu tür hataların belirlenmesi (kestirimi) üzerine kurulmuştur. Bu tür ölçme hataları güvenilirlik için daha önemlidir. Rastlantısal hatalar ne yönde ve miktarda olduğu bilinmeyen hatalardır ve bu özelliğinden hareketle “gerçek puan teorisi” ya da “klasik test teorisi” bir takım sayıtlılar geliştirmiştir. Bu sayıtlıları dikkate alan klasik test teorisi, paralel testlerle yapılan sonsuz sayıda tekrarlı ölçmede gözlenen puanların ortalamasının gerçek puana eşit olacağını öngörür (Crocker ve Algina, 1986). Bu kavramlar klasik test kuramına göre güvenilirlik ile ilgili başlık altında detaylı olarak açıklanmıştır.

Performans düzeyinin belirlenmesi için gerçekleştirilen etkinliğin sonucunun bir şekilde değerlendirilmesi gerekir. Değerlendirmeye konu olan durum; anlaşılır, anlatılabilir, somut ve nesnel olmalıdır. Performansın göstergesi ölçüm işlemi sonucunda belirlenen değerlerin tümüdür (Songur, 1995).

Psikomotor açıdan hedeflenen bir becerinin puanlayıcı tarafından değerlendirilmesinde, oyuncunun, zihin - kas koordinasyonu arasındaki denge algılamasının, davranışı ortaya çıkaran hareketlerin doğruluğunun ve bunları ne kadar sürede yaptığının değerlendirilmesi gerekmektedir. İnsan davranışlarını ölçmede farklılığa neden olabilecek değişkenlerin bilinmesi, değişkenlerin kontrol altına alınabilmesi açısından önem kazanır.

Performans değerlendirme genel anlamda Ataay (1990) tarafından kişinin yeteneklerini, potansiyel güçlerini, davranışlarını ve benzeri niteliklerini diğer kişilerle karşılaştırarak değerlendiren sistematik bir süreç olarak tanımlanmıştır. Nesnel (objektif), tarafsız, sayısal değerlere dayanan bir performans değerlendirmesi, bireyin görevindeki yeterlilik düzeyini ortaya koyabilmek için gereklidir. Bu nedenle, yeterlik ilkesinin gereği olarak yeterlik değerlendirmesi yapılması gerekir. Böyle bir değerlendirme ile bireyin görevine ilişkin güçlü ve zayıf yönleri, performansındaki eksiklik ve aksaklıklar ortaya konabilir (Canman, 2000).

Performans Değerlendirme sistemindeki meydana gelebilecek en temel olumsuzluk, kullanılan değerlendirme ölçeğine bağlı olarak değerlendirmelerin puanlayıcı hatalarından etkilenmesidir. Performans değerlendirme çalışmaları öznel ve çok çeşitli faktörler tarafından etkilenmektedir. Puanlayıcıların, performans değerlendirmelerini adil bir biçimde yürütme ve eksiksiz bir biçimde gerçekleştirmede hata olasılığını artıran faktörleri tanıması bu tür hataları en aza düşürür (Woods, 1997).

Performans değerlendirme, daha önce de belirtildiği gibi performans seviyesinin belirlenmesi için gerçekleştirilen etkinliğin sonucunun (verilerin) bir şekilde değerlendirilmesini gerektirir. İnsana dair değerlendirmelerin de her zaman beraberinde yanlılgı payları taşıdıkları bilinmektedir. Dolayısıyla performans değerlendirme sürecinde bir takım sorunlarla karşılaşmaktadır. Değerlendirme sürecinde karşılaşılan başlıca sorunlar şunlardır (Woods, 1997):

*Ölçme Aracından Kaynaklanan Hatalar:* Değerlendirilecek performansın niteliklerine göre, performans değerlendirmeye yönelik araçlar çeşitlilik gösterir. Değerlendirilen için soyut öğelerden oluşan işlerin değerlendirilmesi oldukça güç olabilmekte iken somut öğelerden oluşması değerlendirmeyi kolaylaştırmaktadır. Örneğin, bir öğretmenin performansını ölçmek soyut verilere dayandığı için oldukça güçken, bir fabrikada makine başında çalışan bir bireyin ürün miktarını ölçme işi somut verilere dayandığı için çok kolaydır. Soyut durumlarda geliştirilen performans değerlendirme araçlarının içeriğini oluşturan sorular hazırlanırken ölçmek istenilen özelliği ne derece

ölçtüğüne ve tekrar ölçme sonucunda aynı değerler alıp almadığı konusunda yani ölçme aracının geçerlik ve güvenilirlik düzeyine çok dikkat etmek gerekmektedir. Ölçme aracının geçerlik ve güvenilirlik düzeyi düşük olması halinde performans değerlendirme sonucu da hatalı olacaktır (Schermerhorn, 1989, s.432 ve Woods, 1997, s.193).

*Aşırı Hoşgörülü ve Aşırı Katı Olmaktan Kaynaklanan Hatalar:* Performans değerlendirmede sık karşılaşılan sorunların başında aşırı hoşgörülü ve aşırı katı olma gelmektedir. İnsan davranışını değerlendirme ve bu davranışları da bir insanın değerlendirmesi söz konusu olmasından ötürü bu tür sorunlarla karşılaşılır. Burada, aşırı hoşgörülü veya aşırı katı olmaktan kaynaklanan bir takım hatalarla birey hak ettiğinden daha az ya da daha çok puan verilerek değerlendirilebilmesine sebep olur (Woods, 1997, s.194).

*Merkezi Eğilim (Standart Ölçme) Hataları:* Standart ölçme hataları, puanlayıcıların değerlendirme esnasında herkesi, performansını göz önünde bulundurmaksızın değerlendirme ölçeğinin orta noktasına yakın bir yerlerde puanlamasıdır. Örneğin bireyin beşli bir ölçekte (1= çok kötü, 5= mükemmel) orta nokta olan 3= orta ile değerlendirilmiş olmasıdır. Bu, bireye yönelik ölçmelerin çoğunda karşılaşılabilecek bir sorundur. Genelde bireye yönelik ölçmelerde kişiyi, ne kadar farklı olursa olsun ortalama veriler çerçevesinde düşünmek, onu ortalamaya yakın görmek, hatalı sonuç verebilmektedir. Standart ölçme hatası, herkesi ortalama veya vasat ölçülerde görme eğiliminden kaynaklanır. Bu nedenle son derece de sakıncalı durumlar doğurur. Çünkü performans değerlendirme sürecinin işletilmesinde önemli bir amaç da kişilerin aralarındaki başarıya yönelik farklılıklarını belirleyebilmektir.

*Yakın Zaman Etkisi Hataları:* Bireyler daha önceden meydana gelen olaylardan çok, en son gerçekleşen olayları hatırlama eğilimi içindedirler. Örneğin bir yöneticinin çalışanı altı ya da sekiz ay önceki performansından daha çok, bir kaç hafta ya da bir ay önceki performansını muhtemelen daha iyi hatırlar. Puanlayıcılar, puanlama süreci boyunca bireylerin performanslarını kayıt etmedikleri sürece büyük bir ihtimalle bireyin en son performansını göz önünde bulundurarak karar verirler (Budak, 1999).

*En Son Ölçme İşlemine Bağlı Kalma:* Puanlayıcıların, bireylerin performansını bir önceki değerlendirmeye benzer bir şekilde değerlendirme eğiliminde olmasından kaynaklanan hatalardır. Örneğin birey en son değerlendirmede yüksek puan almışsa şu andaki performansına bakılmaksızın tekrar yüksek düzeyde puanlanmasıdır (Budak, 1999: 312- 313).

En son ölçmeye bağlı kalma hatasını önlemenin bir yolu olarak puanlayıcılara, puanlama süreci boyunca bireylerin performanslarına ilişkin not tutmaları önerilmektedir. Not tutma, bu tür hataların unutkanlıktan kaynaklananlarını bir ölçüde önleyebilmekteyse de, puanlayıcılar bazen bu hatayı bilinçli olarak yaptıklarını ifade etmektedirler.

*Baskın Özellik (Halo Etkisi):* Puanlayıcının bireyin performansındaki, birbirinden bağımsız ve belirgin özellikler arasındaki farklılıkları görememesi ve bu doğrultuda hatalı değerlendirme yapması Halo Etkisi anlamına gelmektedir (Kaynak, 1998: 222).

Halo etkisinden kaynaklanan hatalar, puanlayıcının bireyin özellikle tek bir pozitif kişilik özelliğini, davranışını ya da eylemini temel alarak değerlendirmesinden kaynaklanır. Bu tür hata, puanlayıcının, bireyin bir alandaki veya konudaki çok gelişmiş özelliğinin etkisinde kalarak bunu kişinin tüm özelliklerine genellemesi ile meydana gelir. Bir anlamda, bireyin çok iyi olduğu bir yönünün etkisinde kalarak tüm performansının çok iyi olduğu yanılgısına düşmek burada söz konusu olabilir. Bireyin bir bütün olarak ele alınması ve değerlendirilmesi, bu hatanın performans değerlendirmede etkili olmasını önlemek için alınması gereken bir önlemdir.

*Tek Yönlü Ölçme:* Değerlendirilen kişinin sadece bir yönden, örneğin üstlerin görüşleri veya müşteri görüşlerinin kullanılması, performans belirlemede tek yönlülüğe neden olur. Kişinin yalnız bir yönünün ve yalnızca bir kişi tarafından değerlendirilmesi ulaşılan sonuçların yanlış olmasına neden olabilir. Performans değerlendirmede amaç bireyi bir bütün olarak ve tüm yönleri ile değerlendirecek şekilde çok yönlü bir bakışı sağlayabilmektir.

*Taraflı Ölçme:* Taraflı ölçme, değerlendirmeyi yanlılığa götürür. Puanlayıcının değerlendirdiği kişiyi sevmesi ya da sevmemesi, kendine yakınlığına göre davranarak bunu değerlendirmeye yansıtması taraflı ölçmeye neden olmaktadır. Aksi takdirde, performans değerlendirme çalışmaları puanlanan bireyler tarafından güvenilir bulunmayabilir ve sisteme duyulan inancın da sarsılması sonucunu doğurabilir.

## **1.1. PSİKOMOTOR BECERİLERİ PUANLAMADA KULLANILAN ARAÇLAR**

Psikomotor becerileri puanlamak için en çok kullanılan araçlar, başlıca işaretleme tipi araçlar ve rubrikler olarak sınıflandırılabilir (Moskal, 2000).

### **1.1.1. İşaretleme Listesi Tipi Araçlar**

#### **1.1.1.1. Kontrol Listeleri**

Kontrol listeleri kullanılarak, gözlenen performansın ya da ürünün belirlenen performans ölçütlerine ne derece uyumlu olduğu belirlenebilir. Bireyden beklenen davranışın özelliklerine ilişkin detaylı bilgileri içeren ve bireyin performansının eksik noktalarını belirleme amacıyla kullanılan araçlar kontrol listeleri olarak adlandırılırlar (Airasian,2001).

Kontrol listelerinde var veya yok, evet veya hayır şeklinde puanlanabilen bir dizi davranış, özellik veya nitelik bulunur. Genellikle, daha küçük parçalara ayrılabilen ve karmaşık davranışları belirlemek için uygun olan kontrol listelerinde örneğin, deney yaparken bir dizi hareketi gerektiren davranışlar açıkça belirtilip sıralanabilir. Bazı değerlendirme listeleri öğrencinin görevi (etkinliği) yerine getirirken sık yaptığı hataları da gösterebilir. Bu durumda +1 gibi bir puan her bir pozitif davranış için , - 1 gibi bir puan her bir hata için, 0 ise davranışın gözlenemediği durumlar için verilirken kontrol listelerinde ‘evet’ ‘hayır’ ; ‘var’ ‘yok’ veya 0-1 şeklinde değerlendirilir. Bu listeler, davranışın gözlemci tarafından gözlemlenme fırsatının olmadığı durumları da belirten

ifadelerin eklenmesiyle davranışın gözlenemediği durumları da belirleme şansı verir (Kubiszyn ve Borich, 2003).

Kontrol listelerinde puanlayıcıya davranışın gözlemlendiğini ya da gözlemlenmediğini ifade edecek sadece iki ölçütün sunulması kontrol listeleri kullanılırken ortaya çıkabilecek olumsuzluklardan biridir. Kontrol listelerinin başka bir olumsuz yanı da, bireyin performansının belli bir puanlama sistemine göre değerlendirilmesidir. Bu durum her ne kadar bireylerin güçlü ve zayıf öğrenme davranışlarını belirlemeye yarasa da, puanlayıcı performansı belli bir puan ile değerlendirmek zorunda kalır ve performans sürecini kaydedemez (Airasian,2001).

Kontrol listelerine göre puanlayıcılar, bireylerin başarılarını değerlendirmekten çok, onların çalışmaları hakkında bir yorumda bulunurlar. Kontrol listesi yönteminin uygulanması, her görev veya görev grupları için farklı liste hazırlamanın zahmetli olması, zaman alması ve değerlendirilmesi oldukça karmaşık olmasından dolayı çok kolay değildir. Bu tip araçlar, gözlenecek davranışın varlığını “x” veya “evet” gibi bir işaret veya kelime ile gösterilmeye imkan veren, fakat derecelenmesine imkan vermeyen araç tipidir. İşaretleme listeleri, genellikle, olumlu veya olumsuz davranışların sayısını kaydetmeye yarayan araçlardır. İşaretleme listesinden elde edilen gözlemlerin toplam puanını çıkarmak da oldukça teknik bir iştir. Ancak dereceleme ölçeğinden toplam puan elde etmek daha kolaydır (Turgut, 1993).

#### **1.1.1.2. Derecelendirme Ölçekleri**

Performansa ilişkin belirli ölçütlerin karşılanıp karşılanmadığıyla ilgili olan sınırlı durumlarda değerlendirme için uygun bir seçim olan kontrol listeleri performans düzeylerini tanımlayan bir yapıya sahip değildir. Derecelendirme ölçekleri ise ölçülen özelliğe ilişkin performansı çeşitli düzeyleriyle tanımlayabilir ve ölçütlerin ne dereceye kadar karşılandığını görmeye olanak sağlar (Moskal, 2000).

Bir kişinin başka bir kişi hakkında sistemli veya sistemsiz gözlem sonucunda edindiği kanıları bildirmesinin diğer bir yolu derecelendirme ölçekleridir. Bu ölçekler

gözlemlerin kaydedilmesi için kullanılan araçlardan biridir. Gözlemlenecek özelliğin öğrencide bulunma sıklığını kaydetmede ve gösterdiği diğer özellikleri hangi seviyelerde gösterdiğini belirlemek amacıyla kullanılır (Oosterhof, 1994). Öğrenciler hakkında öğretmenlerin, sporcular hakkında antrenörlerin, kısaca astlar hakkında üstlerin kanıdır. Derecelendirme ölçekleri bireyin her alandaki özelliği ile ilgili olarak,-örneğin bağımsız davranabilmesi, işbirliği yapabilmesi, liderlik yeteneği, belirgin alışkanlıkları, kendine güveni gibi-çeşitli konularda düzenlenebilir. Derecelendirme ölçekleri, bireyi tanıyan ya da gözleyen herkes tarafından doldurulabilir. Derecelendirme ölçeği formu, bir tür betimleyici sözcükler ya da ifadeler listesi olup, dereceleyecek kimse, gözlediği kişiye uygun bulduklarına işaret koyar. Yapılandırılmış gözleme dayalı ölçmelerde testlerin puanlanması olarak tanımlanan, dereceleme ölçeği gözlenen beceri veya davranışın mükemmelliğini de göstermeye olanak sağlamaktadır. Bu tür ölçeklerde puanlayıcının yargısı önemlidir (Balcı, 2001).

Ölçülen özelliğe ilişkin performans seviyelerinin her birinin ne ifade ettiği hakkında bilgi sunmada yararlıdır. Bireylerin performansları hakkında bilgi edinmenin yanında, bireylerin zayıf ve güçlü oldukları noktalar hakkında bilgide vermektedir. Kullanılan ölçek çeşidine göre 4 çeşit derecelendirme ölçeği bulunmaktadır (Taylor,2003).

- 1- Sayısal derecelendirme ölçeği
- 2- Betimsel derecelendirme ölçeği
- 3- Karşılaştırmalı derecelendirme ölçeği
- 4- Grafik derecelendirme ölçeğidir.

#### **1.1.1.2.1.Sayısal derecelendirme ölçeği:**

Derecelendirilmesi istenen özelliğin genellikle 1 ile 5 arasında derecelendirilmesiyle oluşan ölçeklere verilen addır. Burada kullanılan 1 olumsuz ifadeyi gösterirken 5 olumlu ifadeyi göstermektedir.

Spesifik kategori ölçekleri olarak da isimlendirilen bu tür ölçeklerde puanlayıcı ölçek pozisyonlarınca düzenlenmiş gruplardan birini seçer. Ölçek pozisyon veya gruplarının sayısı, araştırma problemi ve yargıya göre değişmekle beraber tipik olarak 5-11 arası grup bulunur. Grupları gösteren sözel betimlemeler gruba uygun davranışı betimleyen kısa ifadeler şeklinde olmalıdır. Grupların tanımlanmış olması puanlayıcılar arası değişkenliği azaltarak güvenilirliği arttırmaktadır (Johnson ve ark., 2000).

#### **1.1.1.2.2. Betimsel Derecelendirme Ölçeği:**

Derecelendirmesi istenen özelliğin betimsel ifadeler kullanılarak derecelendirilmesi sonucu oluşan ifadelerdir. Farklı performans seviyelerini belirlemek amacıyla farklı tanımlamalar yapılır (Airasian, 2001). Örneğin, bireyin davranışlarına uygun düşenin önündeki parantezin içine bir ( X ) işareti koyarak birey için uygun olan durumu belirtiniz.

- ( ) Ödevlerini hiç yapmaz, sorumluluk duygusu gelişmemiştir.
- ( ) Ödevlerini eksik olarak yapar.
- ( ) Ödevlerini zamanından önce eksiksiz olarak yapar.

#### **1.1.1.2.3. Karşılaştırmalı derecelendirme ölçeği:**

Bu tip ölçeklerde derecelendirilecek özellik, normal olarak diğer bireylerde bulunması beklenen miktar ya da düzey ile karşılaştırılarak bir ölçek üzerinde değerlendirilir. Bir sınıftaki belli bir niteliğe (örneğin liderlik gibi) en üst ve en alt düzeyde sahip olan öğrencileri belirlemek için liderlik özelliğine en az sahip, orta düzeyde sahip ve yüksek düzeyde sahip öğrenciler seçilir. Bundan sonra herhangi bir öğrenci liderlik özelliği yönünden bu üç öğrenci ile karşılaştırılır ve “eşit, daha iyi, daha zayıf” gibi yargılar verilir (Balcı, 2001).

#### **1.1.1.2.4.Grafik derecelendirme ölçeği ( Metrik Ölçek):**

Ölçülecek niteliğin basamaklar halinde yatay bir doğru üzerinde gösterilmesiyle oluşturulmuş ölçeklere verilen addır. Doğru üzerinde bir işaret konularak derecelendirilir. Öğrenci gelişimini ve oluşan ürünlerin değerlendirilmesinde derecelendirme ölçeklerinden faydalanır (Linn, Gronlund, 1995).

Bu tür ölçeklerde puanlayıcı tercih ya da derecesini, özelliğin bir ucundan diğer ucuna uzanan çizgi üzerinde bir işaretle belirtir. Ölçeğin noktaları ya da dereceleri çizgi boyunca kısa açıklamalarla gösterilebilir. Ancak bu nokta ya da derecelerin görevi, puanlayıcıya süresiz kategoriler sağlamak değil onun tercihini yerleştirmesine rehberlik etmektir. Etkili kullanım için uç ifadelerin aşırı olmasından kaçınılmalıdır (Balcı, 2001).

#### **1.1.2. Rubrikler**

Puanlama rubrikleri, bireylerin gösterdikleri performansların sürecini ve sonucunu analiz edebilmek için puanlayıcılar ya da ölçme uzmanları tarafından geliştirilen açıklayıcı puanlama şemalarıdır (Brookhart, 1999). Puanlama öncesi bu tür bir şemanın hazırlanmış olması, puanlamanın öznel olduğu performans değerlendirmede yer alan görev ya da maddelerin daha nesnel puanlanabilmesini sağlar. Rubriklerde her bir puan kategorisine ilişkin cevap özelliklerinin tanımlanmış olması, birbirinden bağımsız puanlayıcıların bir soruya verilen cevaba aynı puanı verme olasılığını arttırır. Bir performans değerlendirmenin puanlanmasında rubriğin uygun bir ölçme tekniği olup olmadığına karar vermek, kimin için kullanılacağından çok yapılan ölçmenin amacına bağlıdır (Moskal, 2000). Rubrikler holistik ve analitik olmak üzere iki sınıfa ayrılmaktadır.

##### **1.1.2.1.Holistik Rubrikler**

1960'larda ortaya çıkan Holistik rubrikler ilk olarak genel izlenimle puanlama yöntemi olarak tanımlanmıştır (Nitko, 2001). Holistik rubrikler bireylerin her bir görev ya da maddede göstermiş oldukları performansın bir bütün olarak, alt bileşenlerine

ayrılmaksızın puanlanmasında kullanılır. Holistik rubrikle yapılan puanlamada, her bir görev ya da maddenin bütününde gösterilen performans yüksek ise bazı bileşenlerindeki hatalar önemsenmeyebilir (Mertler, 2001). Holistik rubrikle puanlamada, puanlayıcı her bir görev ya da maddeyi bir bütün olarak inceleyerek, bu bütüne ilişkin bir puan vermektedir.

### **1.1.2.2. Analitik Rubrikler**

Holistik rubriklerin tam tersine, analitik rubriklerde bireyin her bir görev ya da maddede gösterdiği performans öncelikle alt bileşenlerine ayrılarak puanlanır. Görev ya da maddeye ilişkin puan bu bileşenlerden alınan puanların toplamıdır. Bu analitik rubriklerde puanlamanın çok boyutta yapıldığını göstermektedir (Mertler, 2001). Bir görev ya da maddede gerçekleştirilen performansın her bir alt boyutunda bireyin güçlü ve zayıf yönleri belirlenebilir. Pek çok araştırmacıya göre, bir alanın alt alanlarına ayrılarak yapılan puanlamada, puanlayıcıların öznelliğinden daha az etkilenme olmakta ve puanlayıcılar arası değişkenlik daha azalmaktadır (Alharby, 2006).

Ölçme sonuçlarına karışabilecek hataların kaynaklarına yönelik güvenilirlik hesaplama yöntemleri, ölçme kuramlarının başlıca çalışma alanlarıdır. Klasik test kuramı ve madde tepki kuramının birçok modeli tek bir analizle, tek bir değişkenlik kaynağından gelen hatayı dikkate alır. Oysa birçok ölçme durumunda potansiyel hata ya da değişkenlik kaynağı birden çoktur. Birden çok değişkenlik kaynağından gelen hatanın birlikte değerlendirilmesi ölçmenin uygulama alanının eğitim ve psikoloji alanlarında genişlemenin yanında diğer alanlara da giderek yayılması sonucunda artan bir gereklilik olmuştur.

Sporcuların becerilerini ölçmek için puanlayıcıların ölçme sürecine katılmaları gerekmektedir. Puanlayıcıların, ölçme sürecine katıldıkları ölçme uygulamalarında, güvenilirlik çalışmaları için birden çok değişkenlik kaynağının birlikte değerlendirilerek kullanılmasına ihtiyaç duyulmaktadır. Çoğu zaman ölçme sonuçlarına dayalı olarak önemli kararların verilmesi söz konusudur. Kararların isabetliliği yapılan işlemin karar vermek için uygun olup olmadığı ya da ölçme sonuçlarına karışan hatalardan

arındırılarak bu sonuçların kullanılması ile olanaklıdır. Bu nedenle de ölçme işlemi ve sonuçlarına karışan hatanın araştırılması, yani güvenilirliğin test edilmesi gereklidir.

## **1.2. PERFORMANS DEĞERLENDİRMEDE GÜVENİRLİK**

Ölçme çalışmalarında amaç, ölçülen özellik bakımından elde edilen puanın olabildiğince gerçek puana yakın olmasının sağlanmasıdır (Baykul, 2000). Gerçek puanlara yakın ölçme sonuçları ise ölçme ile elde edilen puanlara karışan hatanın az olması ya da hatasızlığı ile olanaklıdır ki bu da güvenilirlik olarak adlandırılır.

Güvenirlik, ölçme sonuçlarının tesadüfi hatalarından arınık olma derecesi olarak tanımlanmakta olup, bir başka deyişle güvenilirlik, ölçümlerin zaman içerisindeki tutarlılıklarının derecesi olarak adlandırılmaktadır (Baykul, 2000). Güvenirliğin derecesi genellikle bir katsayı ile ifade edilir. Bu katsayısı “0” (güvenilir değil) ile “1”(mükemmel güvenilirlik) arasında değişen değerler alır ve ölçme sonuçlarının tesadüfi hatadan ne derece arınık olduğunu gösterir. Ölçme sonuçları daima bir miktar hata barındırmaktadır, bir başka deyişle ölçme sonuçlarının güvenilirliği daima bir miktar tehdit altında olmaktadır. Genellikle, performansın ölçülmesinde güvenilirlik düşüktür. Ayrıca ürünün ölçülmesi, sürecin ölçülmesine kıyasla daha güvenilirdir. Çünkü, ürün sonradan gözden geçirilebilir; fakat eğer süreç kaydedilmiyorsa ya da kaydedilmemişse, onun tekrar gözden geçirilme olanağı yoktur.

Genel olarak performansın ölçülmesinde farklı değişkenlik kaynakları ortaya çıkar. Bu değişkenlik kaynaklarından biri bireylerdir. Birey değişkenlik kaynağı; bireylerin problem, görev ya da ürünle ilgili bilgi ya da becerisini ortaya koyduğu yetenekleridir. Değerlendirilen bireylerin yeteneklerinin birbirinden farklı olması beklenen bir durumdur. Performansın ölçülmesinde puanlayıcıların oldukça önemli bir yeri vardır ve puanlayıcılar da diğer bir değişkenlik kaynağıdır. Bunun nedeni puanlayıcıların puanlamalarını etkileyen farklı yaklaşım ve tepkilerinin bulunmasıdır. Bir başka değişkenlik kaynağı, bireyden yerine getirmesi beklenen görev ya da maddedir. Performanslara ilişkin görevler ya da maddeler, bireylerin performanslarını gözlenebilir hale getirmelerini sağlayacak kadar açık ve net, aynı zamanda bireyler arasındaki

farklılıkları ortaya çıkarabilecek ayrıntıları içerecek nitelikte olmalıdır. Performansta yer alan görev ya da maddelerin sayısı ve parametreleri bireylerin performansını etkileyen özellikler arasındadır. Performansın ölçülmesinin gerekli olduğu durumlarda ölçmeyi oluşturan bu değişkenlik kaynaklarının ve birbirleriyle etkileşimlerinin incelenmesi gerekir. Bir bireyin ölçülebilir yeteneğinin, gerçekleştirmesi istenilen görevde gösterdiği performans ile kestirilebileceği varsayılır. Ancak bireylerin bu performanslarının gerçekte, puanlayıcıların farklı katılık düzeylerinde ya da maddeler ile görevlerin farklı güçlüklerde olmasıyla etkileşim içinde olduğunu da göz ardı etmemek gerekir (Güler,2008).

Performansın ölçülmesinde puanlamanın puanlayıcıların kararlarına bağlı yapılması yani öznelliği, güvenilirliği düşüren nedenlerden biridir. Puanlayıcı, performans ölçmenin güvenilirliğini düşüren önemli bir hata kaynağı olmakla birlikte, görev ya da zaman gibi diğer faktörlerle olan etkileşimi de en az o kadar önemli bir hata kaynağı olabilir. Bu nedenle sadece puanlayıcılar arası ya da puanlayıcının kendi içindeki, puanlayıcı içi tutarlılığı sağlamak puanlamanın güvenilirliği için yeterli olmamakta, hata kaynakları arasındaki etkileşimin de güvenilirliğin hesaplanmasında göz önünde bulundurulması gerekmektedir. Performans değerlendirmenin güvenilirliği, ölçmenin üç temel kuramı olan klasik test kuramı, madde tepki kuramı ve genellenebilirlik kuramına dayalı yöntemlerle çalışılmaktadır (Güler, 2008).

### **1.3.KLASİK TEST KURAMINA GÖRE GÜVENİRLİK**

Klasik test kuramı gerçek puanı gözlenen puandan kestirmeye çalışılan kuramlardan biridir. Klasik test kuramına gerçek puan modeli de denilmektedir (Crocker ve Algina, 1986). Bir ölçme aracındaki herhangi bir maddeye verilen yanıtlar (X), ölçmeye çalışılan özelliğe bireyin sahip olma düzeyi (T) ve hata terimleri (E) ile ifade edilir. Bu bağıntı klasik test kuramının  $X=T+E$  modeliyle ifade edilir. Eğer tesadüfi hatalardan arınık bir ölçme herhangi  $a$  maddesi için yapıldıysa, bu koşulda ölçme aracından elde edilen puan aynı zamanda bireyin ölçülmek istenilen özelliğe sahip olma seviyesine eşit ( $X=T$ ) olacaktır (Fan, 1998).

Gözlem değeri (X) olan ölçme değeri uygulamalarda ulaşılabilen tek değerdir. Gerçek değer (T) ve hata değeri (E) hakkında bir bilgi elde edilemez. Hataların tesadüfi olarak dağıldığı varsayımıyla, hatalar tesadüfi olarak pozitif ve negatif yönde gelişir ve birbirlerini elemesi sayesinde ölçme hataları ortalaması sıfır olur.(Gürsakal, 2001).

Klasik test kuramında güvenilirlik katsayısı, gerçek puanlar varyansının gözlenen puanlar varyansına oranı veya gözlenen puanlarla gerçek puanlar arasındaki korelasyonun karesi olarak tanımlanır. Gerçek puanlar varyansı ve hata varyansının toplamından ise gözlenen puanların varyansı oluşur. (Lord ve Novick, 1968).

Somut özelliklerin ölçülmesinde kullanılan ölçeklerle yapılan tekrarlı ölçmelerin her defasında birbirine çok yakın sonuçlar alınırken soyut özelliklerin ölçülmesinde kullanılan ölçeklerde ise aynı sonuçların alınması çok zordur. Bu nedenle somut özelliklerin ölçülmesinde kullanılan ölçekler daha güvenilirlerdir. Birimlerin soyut özelliklerinin ölçülmesinde kullanılan ölçekler diğerlerinden bazı önemli ayrılıklar göstermektedir. Birimleri ölçmek üzere, belirlenmiş olan standart aralıkları birbirine eşit olsa bile, başlangıç noktaları keyfi olması nedeniyle bu tür ölçekler miktar anlamı olan ölçmeler verememektedir. Bu sebeple, ölçeğin ölçtüğü özelliğin gerçek değerine yakın ölçmeler verme gücünün titizlikle incelenmesine ve dolayısıyla bu incelemede analitik olmayan yani bir ölçme aracıyla elde edilen ölçmelerdeki değişkenliği aşağıdaki gibi ayırmaya gerek vardır (Özçelik, 1998);

- Ölçülmek istenen özellikteki gerçek farklardan kaynaklanan değişkenlik
- Ölçülmek istenen özelliği ölçülmek istenmeyen özelliklerin etkilemesi sonucunda gerçek farklardan kaynaklanan değişkenlik,
- Yukarıda açıklanan iki değişkenliğin etkileşimden kaynaklanan değişkenlik,
- Ölçme hatasından kaynaklanan değişkenlik.

Klasik test kuramında, güvenilirlik gerçek puan varyansının, gözlenen puan varyansına oranı olarak ifade edilmekte, buradaki gözlenen varyans, gerçek puan varyansı ile hata varyansı toplamına eşit olmaktadır. Elde edilecek ölçmeden ölçmeye değişim

gösterecek olması, hata varyansı ve buna bağlı olarak güvenilirliğin değişeceği sonucunu doğurur. Test tekrar test güvenilirliğinde puanların iki zaman arasındaki değişimi hata olarak düşünülürken, madde örneklemeden ortaya çıkabilecek değişim düşünülmez. İç tutarlılık katsayısı hesaplanırken ise madde örneklemeden dolayı meydana gelen değişim hata kaynağı olarak ele alınırken, bu kez zamana bağlı oluşabilecek değişim hata kaynağı olarak ele alınmaz. Bir başka deyişle, klasik test kuramında güvenilirlik, hata kaynağına bağlı olarak farklılık göstermektedir. Yapılan her bir ölçmenin sonuçlarına göre farklı değişkenlik kaynaklarına ilişkin farklı güvenilirlik katsayıları hesaplanır. Her şeyden önce, hesaplanan güvenilirlik değerleri arasındaki farklılık oldukça fazla olduğunda, hangi güvenilirlik değerinin göz önünde bulundurulacağına karar vermek çok güç olabilir. Örneğin, bir testin test tekrar test güvenilirliğinin yüksek, birinci uygulamanın puanlarına ilişkin iç tutarlılık katsayısının oldukça düşük, ikinci uygulamanın puanlarına ilişkin iç tutarlılık katsayısının ise orta değerler aldığını düşünelim. Bu durumda araştırmacı “hangi güvenilirliğe göre” yorum yapacağına karar vermekte zorlanacaktır (Kieffer, 1998;akt. Güler, 2009).

Paralel formlar yöntemi ile birbirinin paraleli olan ölçme araçlarının birbirleri ile tutarlı sonuçlar verip vermedikleri incelenir. Paralel formlarla elde edilen güvenilirlik katsayısı tutarlılık anlamındadır. Bu yöntemle formların hata kaynağı olması güvenilirlik belirlemede birbirinin paraleli olan formlardan gelen hatalar söz konusu olmasından kaynaklanır. Diğer yandan, iç tutarlılık anlamında güvenilirliğin hesaplanmasında kullanılan yöntemlerde ise maddelerin birbirleriyle ve testin bütünü ile ilişkisi dikkate alınır. Bu nedenle iç tutarlılık güvenilirliğinde maddeler hata kaynağıdır (Traub, 1994).

İç tutarlılık katsayıları, ölçekte, bütün maddelerin aynı özelliği ölçtüğü varsayımına dayanmaktadır. Likert tipi tutum ölçekleri gibi “Tamamen Katılıyorum”, “Katılıyorum”, “Kararsızım”, “Katılmıyorum”, “Hiç Katılmıyorum” gibi derecelendirilmiş ölçme araçlarında Cronbach tarafından 1951 yılında geliştirilen ve Cronbach Alfa katsayısı olarak bilinen formülden yararlanır. Cronbach Alfa iç tutarlılık katsayısının düşük olması ölçeğin birkaç özelliği beraber ölçtüğünü gösterebilir. Çünkü elde edilen alfa değeri, testin homojenliğinin bir göstergesi olarak kabul edilir. Hesaplanan bu iç

tutarlılık katsayısı için Nunnally (1978), genel kabulün en az 0,70 olduğunu belirtmiştir (Tavşancıl, 2002).

Birçok ölçme durumunda puanlayıcılar arası uyum ya da güvenilirliğin belirlenmesi gerekmektedir. Aslında eğitim ve psikoloji gibi sosyal bilimlerin birçok alanında çok sayıda puanlayıcı tarafından yapılan puanlamalar kullanılmakla birlikte, bu alanların dışında da yaygın bir kullanım alanına sahiptir. Bu nedenle puanlayıcılar arası uyum ya da güvenilirliğe ilgi giderek artmakta ve önemli hale gelmektedir.

Yeteneğin ölçülmesinde kullanılan testlerden elde edilen değerlendirmeler için testi alan kişiden, performansını tekrar sergilemesi istenmez. Bunun nedeni performansın yinelenmesi, ikinci performansı etkileyen ve/veya geliştiren bir durumdur. Puanlayıcıların tek bir performansa verdikleri puanlar arasındaki uyum performans ölçmelerinde güvenilirlik, anlamına gelmektedir. Bu puanlar uyumluysa, puanlayıcılar arası güvenilirlik sağlanmış demektir. Aynı performansa farklı puanlayıcılar tarafından verilen puanların tutarlı olması puanlayıcılar arası güvenilirlik anlamını taşır (Mabry, 1999 (Akt: Yıldırım, 1999)).

Performans değerlendirmede, klasik test kuramı yaklaşımına dayalı, çok önemli bir hata kaynağı olan puanlayıcılara bağlı olarak güvenilirliğin hesaplanmasında, puanlayıcı-içi (intrarater) güvenilirlik ve puanlayıcılar-arası (interrater) güvenilirlik hesaplama teknikleri olarak iki tane farklı teknik bulunmaktadır (Alharby, 2006).

*Puanlayıcı içi (intrarater)güvenirlik:* Bir puanlayıcının verdiği puanlardaki kendi kararlılığının göstergesidir. Önceden tahmin edilemeyen bazı geliş-güzel koşullar, örneğin; puanlayıcının psikolojik durumundaki değişmeler, puanlayıcının yorgunluğu kendi verdiği puanlar arasında uyumsuzluğun görülmesine neden olabilir. Bu güvenilirlik, bir puanlayıcının farklı zamanlarda aynı bireylerin performanslarını birden fazla kez puanlaması ile elde edilen puanlar arası korelasyon katsayısının hesaplanmasıyla elde edilir. Test tekrar test güvenilirliğin hesaplanmasına benzemekte olan puanlayıcı içi güvenilirlik yönteminde bir kez uygulanan performans değerlendirmesini bir puanlayıcının, iki farklı zamanda puanlamasıyla elde edilen

puanlar arasındaki korelasyonun hesaplanmasıyla, puanlayıcının puanlamadaki tutarlılığı anlamında güvenilirliğin kestirilmesi söz konusu iken, test tekrar test yönteminde, bir test bir gruba farklı zamanlarda iki kez uygulanır, farklı iki uygulamadan elde edilen puanlar arası korelasyon hesaplanarak, testin zamana karşı kararlılığı anlamında güvenilirliği kestirilir.

Bunun yanı sıra bir puanlayıcının aynı performansları iki kez puanlaması sonucunda elde edilen puanların uyumunun basit yüzdesi tutarlılık anlamında güvenilirliğin bir ölçüsünü vermektedir. Basit yüzde; puanlayıcının aynı puanı verdiği bireylerin sayısının, puan verdiği tüm bireylerin sayısına oranının yüzdesidir. Örneğin; bir puanlayıcının 40 öğrencinin performansını iki farklı zamanda puanladığını düşünüldüğünde, puanlayıcı iki puanlama sonucunda 16 öğrencinin performansına her iki puanlamada aynı puanları vermiş 24 öğrencinin performansına ise farklı puanlar vermiş olsun. Bu durumda puanların uyumunun basit yüzdesi= $100 \times (16/40) = \%40$  olacaktır. Basit yüzde yönteminin en büyük avantajı, hesaplanmasının kolay ve anlaşılır olmasıdır. Diğer taraftan, sınıflama, sıralama, aralık ve oran ölçeğinde elde edilen verilerin tümüne uygulanabilme özelliğine sahiptir (Goodwin, 2001). Tesadüfen veya şansla ortaya çıkan uyumu hesaplayamaması basit yüzde hesaplanmasının en önemli sınırlılığı olarak görülmektedir.

*Puanlayıcılar-arası (interrater) güvenilirlik:* Klasik Test Kuramında farklı puanlayıcıların birbirinden bağımsız olarak her bir bireyin performansına aynı puanı verme tutarlılıklarının derecesidir. Puanlayıcılar-arası güvenilirlik en yaygın olarak, iki puanlayıcının performansa verdikleri puanlar arasındaki doğrusal ilişkinin derecesi olan korelasyon katsayısı (güvenirlik katsayısı olarak da bilinir) -Pearson çarpım moment korelasyon katsayısı (r)- ile hesaplanır. Bu korelasyon katsayısı; gerçek varyansın toplam varyansdaki yüzdesi olarak elde edilir. Burada geriye kalan yüzde puanlayıcıların puanları arasındaki uyumsuzluğu gösteren hata anlamı taşımaktadır (Goodwin, 2001).

Puanlayıcılar arası güvenilirliği istatistiksel olarak değerlendirebilmesi açısından uyum yüzdesi, korelasyon teknikleri, Cohen'in kappası, ağırlıklandırılmış kappa vb. yöntemler bulunmaktadır. Bu araştırma kapsamında puanlayıcılar arası güvenilirliği Kendall'ın uyuşum katsayısı tekniği ile incelenmesi tercih edilmiştir.

Kendall uyuşum katsayısı (W), çok sayıda puanlayıcının vermiş olduğu puanlar arasında ne ölçüde uyuşma bulunduğunu belirler. Puanlayıcıların verdikleri puanlar sıralı ölçek verisi niteliğinde olmaktadır (Howell,2002).

İkiden fazla puanlayıcı, kişilerin davranışlarını gözlemliyorsa, bu durumda puanlayıcılar arası güvenirlığın saptanmasına ihtiyaç duyulabilmektedir. Bu durumda ise puanlayıcılar arasındaki tutarlılığı saptamak için, parametrik olmayan istatistiksel bir teknik olan Kendall'ın Konkordans (Kendall'ın uyuşum) katsayısı kullanılmaktadır (Martin ve Bateson, 1992, akt. Tavşancıl, 2002). Puanlayıcılardan alınan ağırlıklı seçenek puanları arasında uyum olup olmadığını test eden Kendall'ın uyuşum katsayısı, n kişinin veya nesnenin ikiden fazla sıralanışı arasındaki bağıntının derecesini veren bir korelasyon ölçüsüdür. Özellikle yargılar arası veya testler arası güvenirlilik araştırmalarında kullanılmakta olan bu ölçü, sıralanabilir veriler için puanlayıcılar arası güvenirlilik katsayısı olarak tanımlanmaktadır (Howell,2002).

İkiden daha fazla değişken söz konusu olduğunda puanlayıcılar arası korelasyonu ölçmek için kullanılan Kendall'ın uyuşum katsayısı testi; k gözlem ya da sıraların kümesi arasındaki uyumu ölçmeye yöneliktir. 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 dayalı olarak sıralayabilir. Bu tür durumlarda kullanılan Kendall uyum katsayılarının varsayımları aşağıdaki gibidir (Kıroğlu, 2001).

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

Kendall'ın uyuşum katsayısı kullanılarak yapılan değerlendirmelerde puanlayıcılar arasındaki uyuşmanın en az 0,80 seviyesinde olması istenmektedir (Szymanski ve Linkowski, 1993).

Klasik test kuramında, tüm yetenek seviyelerindeki bireyler için ölçme hatalarının aynı olduğu varsayılmaktadır. Bu nedenle klasik test kuramının bir sınırlılığı ölçme hatalarına ilişkin varyansın bütün bireyler için aynı kalmasıdır. Bazı bireyler bir testte diğer bireylerden daha tutarlı davranabilir ve bu tutarlılık yetenek düzeyi ile birlikte değişme gösterdiğinden yüksek yetenek düzeyine sahip bireyler, testlerin paralel formlarında, orta yetenek düzeyine sahip bireylerden daha tutarlı sonuçlar elde ederler. Bu da klasik test kuramına dayalı olarak geliştirilen ölçme araçlarının test puanlarına ilişkin verdiği bilginin ve psikolojik özellikleri kestirmedeki gücünün yetersiz ve eksik olmasından kaynaklanmaktadır (Hambelton ve Swaminathan, 1985).

Klasik test kuramının bu varsayımlardan kaynaklanan sınırlılıkları incelendiğinde aşağıdaki kavramlardan söz edilmektedir (Crocker ve Algina, 1986):

- Güvenirlik kavramının paralel formlarla tanımlanmış olması, güvenirlik hesaplamalarında, kaynağı belli olmayan tesadüfi hatalardan dolayı, güvenirliliğin alt sınırı elde edilir.
- Madde ayırıcılık ve madde güçlük indekslerinin değeri, testin uygulandığı cevaplayıcı grubunun yeteneğine bağlı olmaktadır.
- Ölçme hatalarının varyansı bütün bireyler için aynı kalmaktadır
- Testler, orta yetenek düzeyindeki gruplar için geliştirilmiştir. Bunun sonucu olarak, yüksek ve düşük yetenek düzeyindeki gruplar için doğru yetenek hesaplamaları elde edilememektedir.
- Bir testin güçlüğü, bireyin yetenek düzeyi ile eşlendiğinde, belli bir test maddesini doğru cevaplama olasılığının ne olacağı kestirilemez. Özellikle, bir veya daha fazla bireyin olduğu grubun test puanlarının özelliklerini gözlemek isteyen test geliştiricisi için bu durum önemlidir.

- Klasik test kuramı, yanlış maddelerin belirlenmesi, test geliştirme, maddelerin maksimum ayırıcılık gücü, test puanlarının eşitlenmesi ve ayırdıkları yetenek düzeyi hakkında bilgi vermekte yetersiz kalmaktadır.

Klasik test kuramının gruba bağımlı istatistikler ortaya koyması ve buna bağlı ortaya çıkan sınırlılıklar yeni kuram arayışlarına sebep olmuştur (Crocker ve Algina, 1986:s.339).

#### **1.4.GENELLENEBİLİRLİK KURAMINA GÖRE GÜVENİRLİK**

Temeli varyans analizine (ANOVA) dayalı istatistiksel bir kuram olan Genellenebilirlik Kuramı, davranış ölçmede güvenilir gözlemlerin tasarlanmasını ve kuramsallaşmasını, güvenilirliğin değerlendirilmesini sağlar. Bu kuram Cronbach ve arkadaşları (1963) tarafından, klasik güvenilirlik kuramının günümüzde hala çok geçerli olan gerçek puan modelinin sınırlılıklarına olan tepkiler doğrultusunda ortaya atılmıştır. Klasik güvenilirlik kuramı, bir tek gerçek puana sahip her bir gözlem ya da test puanının paralel gözlemlerin bir grubuna ait tek bir güvenilirlik katsayısı üretmesi fikri etrafında toplanır (Lord ve Novick, 1968; Baykul, 2000).

Klasik test kuramı, paralel formlar dikkatli bir şekilde eşitlendiğinde elverişli olmasına karşın, örneğin puanlayıcıların merkezi eğilim ve varyansları farklı, gözlemlerden oluşan kapsama bağlı ve yapılar açık olarak ayrışık olduğunda elverişli olmaz ve oldukça sınırlayıcıdır. Birkaç yazar 1950-1960 yılları arasında, aynı gözlemlerin paralel testlerin bir setinden daha çoğuna ait olduğu düşünüleceği ve bu nedenle birden çok güvenilirlik katsayısına sahip olduklarını tartışmışlardır (Cronbach, Rajaratnam ve Glaser, 1963).

Genellenebilirlik kuramında klasik test kuramının aksine, aynı anda birden fazla hata kaynağı bir arada göz önünde bulundurularak tek bir güvenilirlik değerine ulaşmak mümkün olmaktadır. Genellenebilirlik kuramı, sonuçlarının belirli bir örneklemeden çok daha geniş olan evrenine nasıl genellenebileceği ile ilgilenirken belirli bir ölçme sonucuna ya da gözlenen puana odaklanmaz. Ölçülmek istenilen özelliğe ilişkin tüm

ölçme sonuçlarını kapsayan evren genellenebilirlik evrenidir. Bir puanın genellenebilirlik evreni adı verilen daha geniş durumlara doğru bir şekilde nasıl genellenebileceğine bağlı olan o puanın kullanışlılığıdır. (Güler, 2009).

Genellenebilirlik kuramı, kavram gruplarını açıklayarak, güvenilirliğin bir ya da benzer bütün özelliklerini tahmin için gereken yöntemleri sunar. Bunun ötesinde uygulanabilir. Bir genellenebilme analizini yürütmek için; potansiyel çok kaynaklı hataların etkilerini açıklamak mümkünken aynı zamanda tasarımın gelişmesi için gerekli hatayı azaltma ve ölçmenin kesinliğini artırmaya karar vermek de mümkündür.

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).

Genellenebilirlik Kuramı, sadece Klasik Kuramın güvenilirlik kuramına karşı çıkmaz. Aynı zamanda geleneksel olan güvenilirlik ve geçerlik ayrımının yapılmasını da gösterir. Geleneksel olarak geçerliği oluşturma, gözlemlenen ölçümlerin gizli kalmış yapılarına dayanarak alınan çıkarımlar olarak düşünülür. Öyleyse güvenilirlik de, paralel ölçmeler boyunca gözlemlerden alınan gerçek puanlardan yapılan çıkarımlar olarak düşünülmektedir. Klasik kuramda gerçek puanlar, Cronbach ve beraberindeki çalışanlar tarafından genellenebilirlik kuramında evrensel puanlar olarak değiştirilmiş ve tanımı da şu şekilde verilmiştir: Amaçlanan değişkenlik kaynakları için kabul edilebilir gözlemlerin evreninden elde edilecek puanların ortalaması olarak bulunan ölçme puanıdır. Araştırmacılar gözlemlenen puanları, evrensel puanların bir fonksiyonu olarak görür. Bu da o kişinin evrenin bir örneği olarak genellenmesini sağlar. Güvenirlik sorusu ise bu açıdan genellenebilirliğin ne kadar geçerli olduğuna bağlı olarak açıklanır (Cronbach, 1984).

Aynı ölçmeye ilişkin olarak klasik test kuramının farklı hata kaynaklarını dikkate alan farklı yöntemlerle hesaplanan güvenilirlik katsayıları birbirinden farklı olabilir. Genellenebilirlik kuramı klasik test kuramının bu sınırlılığını ortadan kaldırmaya yönelik esnek bir alternatif olarak, puanlayıcı, zaman, ölçme aracı, madde ve görev gibi

bütün potansiyel deęişkenlik kaynaklarından gelen potansiyel hataları birlikte ve eş zamanlı deęerlendirerek kapsamlı tek bir güvenilirlik katsayısı hesaplanmasını sağlar. Genellenebilirlik kuramında bireyin bir performans deęerlendirmede yer alan görevde gösterdiği performansı, puanlayıcı ve görev gibi olası tüm deęişkenlik kaynaklarının bir arada bulunduğu karmaşık bir evrendeki performansından çekilen bir örnekleme olarak görülür. Shavelson ve Webb (1991)'e göre, G kuramı dört farklı açıdan klasik test kuramının daha genişletilmiş bir halidir:

- Genellenebilirlik kuramında çoklu varyans kaynakları tek bir analizde ele alınır.
- Genellenebilirlik kuramı, her bir varyans kaynağının büyüklüğünün belirlenmesini sağlar.
- Bireylerin performanslarına dayalı göreceli kararlar alırken aynı zamanda bireylerin performanslarıyla ilgili mutlak kararlar alınmasına ilişkin güvenilirlik katsayısının hesaplanmasına imkan verir.
- Belirli bir amaca baęlı olarak, ölçme hatasının en aza indirgenebileceği ölçmelerin düzenlenmesine (D-çalışmaları) imkan tanır.Özetle, Genellenebilirlik Kuramı performans deęerlendirmenin güvenilirliğinin kestirilmesine uygun bir kuramdır.

Bazı kavram ve terminolojilerin açıklanması genellenebilirlik kuramını açıklarken konunun daha iyi anlaşılmasında yararlı olacaktır. Genellenebilirlik kuramında deęişkenlik kaynağı (facet), madde ya da görev, zaman, puanlayıcı gibi ölçmenin benzer durumlarının setine verilen isimdir (Brennan, 2001). Dięer bir ifadeyle, ölçme hatasının olası kaynağı deęişkenlik kaynağı olarak tanımlanabilir. Bu tür bir varyansın olabildiğince az olması istenen bir durumken, deęişkenlik kaynağıyla ilişkili bir varyansın yüksek olması istenmeyen bir durumdur (Alharby, 2006). Koşullar (conditions) ise deęişkenlik kaynaklarının düzeyleri olarak açıklanabilir. Örneğin; “görevler” bir deęişkenlik kaynağı ise birinci görev, ikinci görev, üçüncü görev vb.nin her biri bir koşuldur. Var olan bir deęişkenlik kaynağının olası koşullarının genellikle

sonsuz büyüklükte olduğu varsayılır. Gözlemlerin yapıldığı eldeki örneklemin yerine geçebilecek olası gözlemlerin tümüne “kabul edilebilir gözlemlerin evreni” denir. Araştırmacının genellemek istediği koşulların setine ise “Genelleme evreni” adı verilir. Birçok değerlendirme durumunda ölçmenin amacı bireylerdir. Diğer bir ifadeyle istenilen kararların alınacağı ölçmenin hedefleri bireylerdir. Bu nedenle bireylere bağlı varyans istenilen bir durum olup bireyler değişkenlik kaynağı olarak düşünülmez. Amaçlanan değişkenlik kaynakları için kabul edilebilir gözlemlerin evreninden elde edilecek puanların ortalaması olarak bulunan, ölçme puanı Evren puanı olarak adlandırılır. Evren puanı varyansı, klasik test kuramında yer alan gerçek puan varyansına benzer. Genellenebilirlik kuramında iki ayrı hata varyansı bulunması klasik test kuramından farklılıklarından biridir. Genellenebilirlik kuramında iki ayrı anlamda karar vermenin mümkün olması bu farklılığın kaynağıdır. Genellenebilirlik kuramında hem bağıl hem de mutlak hata varyansı bulunmaktadır. Genellenebilirlik kuramındaki göreceli hata varyansı klasik kuramdaki hata varyansı gibi düşünülebilir (Shavelson ve Webb, 1991). Özetle klasik test kuramında;

$$\text{Güvenirlilik katsayısı} = \frac{\text{Gerçek Puan Varyansı}}{\text{Gerçek Puan Varyansı} + \text{Hata Puanı Varyansı}}$$

Genellenebilirlik kuramında;

$$\text{Genellenebilirlik katsayısı} = \frac{\text{Evren puan varyansı}}{\text{Evren Puan Varyansı} + \text{Bağıl Hata Puanı Varyansı}}$$

olarak ifade edilebilir.

Genellenebilirlik kuramında, temel kavram ve terimlere ilaveten farklı desen ve yöntemler de bulunmaktadır;

1. Genellenebilirlik (G) çalışmaları(G-Study),
2. Karar (D) çalışmaları (D-Study),
3. Bağıl ve mutlak hatanın hesaplanmasının mümkün olması ve sonuçların yorumlanmasının bağıl ya da mutlak anlamda yapıyor olması,

4. Genellenebilirlik (G) katsayısının ya da güvenilirlik (Phi) katsayısının hesaplanmasının mümkün olması,
5. Değişkenlik kaynağının tesadüfi (random) ya da sabit (fixed) olması,
6. Verilerin toplanmasında desenin çapraz (cross) ya da yuvalanmış (nested) olması.

Bunların her biri aşağıda kısaca açıklanmıştır.

#### **1.4.1.Genellenebilirlik (G) çalışmaları (G-Study)**

Farklı koşullardaki ölçmelerden kaynaklanan çeşitli kaynakların ölçme hatalarını araştırmak ve değişkenlik kaynaklarına ait varyans bileşenlerini hesaplamak için tasarlanan çalışma G çalışmasıdır. Bir G çalışmasının amacı çeşitli kaynaklardaki gözlenen puanların varyansını tespit etmek ve bu varyans bileşenlerini ayrıştırmaktır (Rentz, 1987; akt: Nalbantoğlu, 2009). G çalışmaları mümkün olduğunca en kapsamlı şekilde, kabul edilebilir gözlemler evrenini tanımlamalıdır (Shavelson&Webb, 1991).

G çalışması davranış ölçmelerinin güvenilirliğinin değerlendirilmesi ve ölçme hatalarını belirli değişkenlik kaynaklarına ayırmak için düzenlenir. G çalışması düzenlemede ilk olarak potansiyel değişkenlik kaynaklarının (madde ya da görev, puanlayıcı, zaman vb.) ve ölçme örnekleminin genelleneceği evrenin belirlenmesi gereklidir. Genellemenin evreni karar vericilerin genellemek istedikleri koşulların seti olarak tanımlanır (Lord ve Novic 1968).

Genellenebilirlik kuramı içinde tanımlanan hata puanları varyansı ise, çeşitli hata kaynakları ve bunların etkileşiminden gelen hata bileşenleri ile kalan hata varyansından oluşur. Bu varyans bileşenleri varyans analizi ile hesaplanabilir. G çalışması önceden tanımlanmış çok faktörlü değişkenlik kaynaklı bir evrende, bu faktörlerin gözlenen puanlar üzerindeki etkisini ve faktörlerin gözlenen puanların varyansına katkısı araştırır. Eldeki ölçme sonuçlarının var olan durumuna göre bir güvenilirlik kestiren çalışma G çalışmasıdır (Brennan, 2001; Shavelson ve Webb, 1991).

G-çalışması sürecinde, örneklemin evrene genellenebilmesi için, puanların değişkenliğinin tüm kaynakları (varyans bileşenleri) ve bunlar arasındaki etkileşimler aynı anda ANOVA yöntemi kullanılarak kestirilir. Kestirilen bu varyans bileşenleri bir sonraki aşama olan D-çalışmasında kullanılır (Güler,2008).

#### **1.4.2.Karar (D) çalışmaları (D-Study)**

D Çalışması, G çalışmasından elde edilen bilgileri kullanarak belli bir amaçla yapılan bir ölçmedeki hataları en aza indirmenin yollarını araştırmak için düzenlenir (Brennan 2001; Crocker ve Aİgina 1986; Shavelson ve Webb 1991). Bir D çalışması ile araştırmada kullanılan değişkenlere ait değerler değiştirilerek hatayı en aza indirmek ve güvenilirliği maksimum seviyeye çıkartmak için çeşitli tasarımlar yapmak mümkündür. Yapılan bir D çalışmasında, incelenen bireyleri tanımlamak için veri toplanabilir. Örneğin, bireyler arasında seçme yapmak ya da bireyleri bir programa yerleştirmek, bir sınav durumundaki grupları karşılaştırmak ya da iki ya da daha fazla değişken arasındaki ilişkiyi araştırmak için D-çalışması düzenlenir (Alharby, 2006). D-çalışması sürecinde araştırmacı, çeşitli senaryolar kullanarak daha yüksek güvenilirlik ve daha düşük hataların olduğu durumlar elde etmeye çalışır. Örneğin; “Daha fazla sayıda görev kullandığımda ne olur?”, “Puanlayıcı sayısını azaltırsam ölçme aracımın güvenilirliğinde kaybım çok fazla olur mu?” vb. sorularının cevabı D-çalışması yapılarak araştırılır. Özetle, pek çok D-çalışması tek bir G-çalışmasından elde edilen aynı varyans kestirimlerine dayalı olarak düzenlenebilir. Diğer bir ifadeyle D-çalışması, ölçme durumlarında ölçme aracında yer alan maddelerin, ölçmenin uygulanma zamanlarının ya da ölçmeyi yapan puanlayıcıların sayılarının iki, üç... katına çıkarıldığında ya da belli oranlarda azaltıldığında ne olacağını kestirmede kullanılabilir. D-çalışmasında kullanılan formül Spermen-Brown formülüne benzerdir (Mushquash ve O’Connor, 2006;akt: Güler, 2008). Karar üzerine temellenen verileri toplamak için düzenlenen D çalışmasının amacı, ölçme ile elde edilmek istenilen evren puanının yeterince güvenilir bir kestiriminin sağlanması için gözlemlerin evrenin farklı değişkenlik kaynaklarının sayısını yeterince örnekleyen bir ölçme düzenlemektir (Brennan, 2001).

Karar çalışmasında genellenebilirlik katsayısı (G) ve Phi ( $\Phi$ ) katsayısı olmak üzere iki tür güvenilirlik katsayısı hesaplanabilir. Fakat hesaplanan bu güvenilirlik katsayıları hata varyansına göre değişiklik göstermektedir. Bu hata varyansları mutlak (absolute) hata ve bağıl (relative) hata varyansı olarak adlandırılmaktadır.

### 1.4.3.Bağıl Hata ve Mutlak Hata

Mutlak hata, bir kişinin gözlenen puanı ve onun evren puanı arasındaki fark olarak tanımlanır. Mutlak hata varyansı  $\sigma^2(\Delta)$  şeklinde gösterilir ve bireylerdeki değişim gruba bağlı olmadan mutlak bir ölçüte göre belirlenir.

Bağıl hata, gözlenen sapma puanı ve evren sapma puanı arasındaki farklılıktır. Bağıl hata varyansı  $\sigma^2(\delta)$  şekilde gösterilir ve klasik test kuramındaki hata varyansına karşılık gelir (Brennan, 2001). Bağıl hata varyansında bireylerin başarıları içinde bulunduğu gruptaki puanların sıralamasına bağlıdır (Rentz,1987;akt:Nalbantoğlu, 2009 ).

Genellenebilirlik kuramı davranış ölçmelerinde bağıl (relative ya da norm-reference) ve mutlak (absolute ya da criterion reference) değerlendirme olmak üzere iki tür değerlendirmenin söz konusu olduğunu dikkate alır (Brennan, 2001; Goodwin, 2001).Genellenebilirlik kuramında bir ölçümün genellemesinin nasıl yapılabileceği karar çalışmasında verilerin nasıl kullanılacağına bağlıdır. Sosyal bilimlerdeki ölçmeler genel olarak bireyleri ya da grupları sıralamak veya bireylerin bilgi, beceri, tutum, yetenek düzeyini hesaplamak için kullanılmaktadır. Bireyler birbiriyle kıyaslanacaksa, her bir bireyin bir göreve ilişkin göstermiş olduğu performansın puanlanması, içinde bulunduğu gruptaki diğer bireylerin nasıl performans gösterdiklerine bağlı olarak yapılacaktır (bu bireyin grup içindeki sıralaması nedir?). Bu durumda araştırmacı verilerin analizi için bağıl (göreceli) modeli kullanacaktır. Bağıl genellenebilirlik katsayısı, bireylerin ham puanlarının büyüklüğündeki olası değişimlere bakılmaksızın, değişkenlik kaynaklarındaki sıralamadaki yerlerini koruma derecelerini göstermektedir. Bağıl genellenebilirlik katsayısı, klasik test kuramındaki güvenilirlik katsayısına karşılık gelmektedir (Mushquash ve O'Connor, 2006).

Mutlak genellenebilirlik katsayısı, bağıl genellenebilirlik katsayısından daha katıdır ve ölçülen bireylerin sıralamadaki yerlerinin düzenindeki uyumun derecesini yansıtmamasının yanı sıra ham puanlarının büyüklüklerindeki uyumun derecesini de yansıtır. Elde edilen puanların gerçek değerlerinin önemli ya da daha anlamlı olduğunda mutlak genellenebilirlik katsayısı, araştırmacı için kullanışlı olabilecek bir katsayıdır (Mushquash ve O'Connor, 2006).

#### **1.4.4.Genellenebilirlik Katsayısı (G-katsayısı) ve Güvenirlik İndeksi (Phi-katsayısı)**

Genellenebilirlik katsayısı (G-katsayısı), klasik test kuramındaki gerçek varyansın gözlenen varyansa oranı olan, güvenirlik katsayısının benzeridir. Klasik test kuramındaki gerçek puan Genellenebilirlik kuramında evren puanı olarak adlandırılır. Bağıl modellerde kullanılır ve gerçek varyansın, gerçek varyans ile bağıl varyansın toplamına oranı G-katsayısını verir. Diğer bir ifadeyle, genellenebilirlik katsayısı, evren puanının varyansının gözlenen puan varyansına oranıdır. Evren puanının varyansı ve bağıl hata varyansının ( $\sigma^2(\delta)$ ) toplamı gözlenen puan varyansına eşittir (Shavelson ve Webb, 1991; Brennan, 2001). Phi ( $\Phi$ ) katsayısı mutlak değerlendirmeler için uygundur ve hata varyansı olarak mutlak hata varyansı ( $\sigma^2(\Delta)$ ) kullanılır. Evren puanının varyansının, evren puanının varyansı ve mutlak hata varyansının toplamına oranı Phi ( $\Phi$ ) katsayısını verir. Belirli bir kesme puanının üzerindeki puanın önem taşıdığı performans ölçümlerinde mutlak G katsayısı tercih edilebilir (Lee ve Frisbie, 1999; Brennan, 2001). Elde edilen puanların puan sıralamasındaki yerinin önem taşıdığı durumlarda bağıl G katsayısını kullanmak uygun olacaktır. Bağıl ve mutlak kararlar için hesaplanan G katsayılarındaki karışıklığı ortadan kaldırmak üzere, bağıl kararlar için hesaplanan değere G katsayısı, mutlak kararlar için hesaplanan değere phi ( $\Phi$ ) katsayısı ya da güvenirlik (dependability) katsayısı adı verilmektedir.

#### **1.4.5. Tesadüfi Değişkenlik Kaynağı ve Sabit Değişkenlik Kaynağı**

G kuramı hem sistematik hem de sistematik olmayan hata kaynaklarını ele alır ve aynı zamanda hesaplar. G kuramı yaklaşımı ile güvenirlik kestirmenin en önemli avantajı,

bütün deęişkenlik kaynaklarını dikkate alarak, kapsamlı bir güvenilirlik elde edilmesidir. G kuramında, bir tek çalışmayla potansiyel bütün hata kaynaklarının ayrı ayrı ve birbiriyle etkileşiminden meydana gelen hatalar değerlendirilerek bir katsayı elde edilir (Brennan, 2001). Genellenebilirlik kuramında araştırmacının kararına baęlı olarak deęişkenlik kaynağının tesadüfi olarak mı yoksa sabit olarak mı ele alınacağına karar verilir. Eęer araştırmacı örneklemin ötesinde bir genellemeye gitmek istemiyorsa, deęişkenlik kaynağı sabit olarak düşünülecektir, öte yandan eęer araştırmacı, örneklemin ötesinde genelleme evrenine genelleme yapmak istiyorsa, bu durumda deęişkenlik kaynağı tesadüfi olarak ele alınacaktır.(Alharby, 2006).

Araştırmacı örneklemin ötesinde genellemeye gitmek istemedięi düşünöldüğünde, örneğin bir test; sosyal, matematik, fen gibi alt testlerden oluşuyorsa koşulların deęiştirebilmek söz konusu olmamakla birlikte alt deęişkenlik kaynaklarının genellenebilirlik evrenindeki koşulları aynıdır; bu durumda deęişkenlik kaynağı sabittir (Shavelson&Webb, 1991).

Araştırmacının örneklemin ötesinde genelleme evrenine genelleme yapmak istedięi düşünöldüğünde eęer bir örneklemin boyutu evrenin boyutundan çok daha küçükse örneklem random olarak kabul edilebilir ve bu örnekte evrenden çekilebilecek aynı boyutta başka bir örneklem deęiştirilebilir olarak kabul edilir çünkü evrenden rastgele olarak çekilmiştir. Bir deęişkenlik kaynağı random olarak davranıyorsa “Bir örneklemedeki koşulları evrendeki aynı boyuttaki başka bir koşul grubuyla deęiştirmek isteyebilir miyim?” sorusu sorulmalıdır. Örneğin bir uygulayıcı matematik başarı testindeki 50 soruyu başka 50 madde örneklemiyle deęiştirmek istediğinde deęişkenlik kaynağı tesadüfidir (Shavelson&Webb,1991).

#### **1.4.6. Çaprazlanmış Desen ve Yuvalanmış Desen**

Verilerin çaprazlanmış ya da yuvalanmış olmaları da Genellenebilirlik kuramındaki desenlerin belirlenmesinde göz önünde tutulmaktadır. Verilerin çaprazlanmış olması, bir deęişkenlik kaynağının bütün koşullarının dięer bir deęişkenlik kaynağının tüm koşullarında gözlenmesini ifade eder ve iki kaynak arasına “ x ” simgesi konularak gösterilir. Verilerin yuvalanmış olması ise, bir deęişkenlik kaynağının bütün koşulları

diğer bir deęişkenlik kaynaęının sadece bazı koşullarınca gözlemlenmesidir ve iki kaynak arasına “:” işareti konularak gösterilir (Shavelson ve Webb, 1991; Brennan, 2001). Bu açıklamalara örnek olarak, eđer her bir sporcu (b) her bir görevi (g) yapar ve her bir puanlayıcı (p) da her bir sporcunun her görevini puanlarsa, bu desen tümüyle çaprazlanmış bir desen olacaktır. Bu desen “b x g x p” şeklinde gösterilir. Çaprazlanmış desenler mümkün olan tüm deęişkenlik kaynaklarından ve bunlar arasındaki etkileşimlerden ortaya çıkan hataların kestirimlerine olanak sağladığından dolayı araştırmacılar tarafından tercih edilirler (Mushquash ve O’Connor, 2006). Çaprazlanmış desenlerin dışında, eđer farklı sporculara (b), farklı görevler (g) sunulursa ve farklı puanlayıcılar (p) farklı sporcular tarafından yapılan farklı görevleri puanlarsa, bu desenlere de tümüyle yuvalanmış desen adı verilir. Bu desende puanlayıcılar görevlerle, görevler de sporcularla yuvalanmıştır ve bu desen “p : g : b” olarak ifade edilir. Benzer şekilde bazı deęişkenlik kaynaklarının çaprazlanmış bazılarının ise yuvalanmış olduđu karışık desenler de düzenlenebilir. Sporcuların her birinin performansı farklı puanlayıcılar tarafından gözlemleniyorsa g x (b:p) ile şeklinde ifade edilir (Atılğan, 2004).

#### 1.4.7. Deęişkenlik Kaynaęı olarak Puanlayıcılar

Varsayımsal bir örnekle “puanlayıcılar” ın genellenebilirlik kuramında bir deęişkenlik kaynaęı olarak nasıl ele alınacağı açıklanmaya çalışılacaktır. Deęişkenlik kaynaęı olarak ele alınan “puanlayıcılar” ın durumundaki deęişiklik bu örneğe dayanarak, ilgili hata terimlerinin de deęişmesine sebep olacaktır. Görevlerin ve puanlayıcıların iki deęişkenlik kaynaęını oluşturduđu, sporcuların ise ölçmenin amacı olduđu bir voleybol becerileri performans deęerlendirmesi ele alınacaktır. Buradan bir G-çalışması ve farklı D-çalışması senaryoları oluşturulacaktır. Buradaki G-çalışması iki deęişkenlik kaynaklı tamamıyla çaprazlanmış (b x g x p) bir desen üzerinden yapılmaktadır. Bu durumda, toplam gözlenen varyans aşağıda ifade edilen yedi bağımsız varyans bileşeninden oluşur (Alharby, 2006):

$$\sigma^2(x_{bgp}) = \sigma^2(b) + \sigma^2(g) + \sigma^2(p) + \sigma^2(bg) + \sigma^2(bp) + \sigma^2(gp) + \sigma^2(bgp) \quad (1)$$

Her bir varyans bileşeni, varyans analizi ile aşağıda verilen Tablo:1 yardımıyla kestirilebilir:

**Tablo:1 İki Değişkenlik Kaynaklı Tesadüfi Desen için Varyans Bileşenlerinin Kestirilmesi**

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	Kestirilen Varyans Bileşenleri
Birey (b)	SS <sub>b</sub>	n <sub>b</sub> -1	MS <sub>b</sub> = SS <sub>b</sub> / n <sub>b</sub> -1	$\sigma_{(b)}^2$
Puanlayıcı (p)	SS <sub>p</sub>	n <sub>p</sub> -1	MS <sub>p</sub> = SS <sub>p</sub> / n <sub>p</sub> -1	$\sigma_{(p)}^2$
Görev (g)	SS <sub>g</sub>	n <sub>g</sub> -1	MS <sub>g</sub> = SS <sub>g</sub> / n <sub>g</sub> -1	$\sigma_{(g)}^2$
b x p	SS <sub>bp</sub>	(n <sub>b</sub> -1)( n <sub>p</sub> -1)	MS <sub>bp</sub> = SS <sub>bp</sub> / n <sub>bp</sub> -1	$\sigma_{(bp)}^2$
b x g	SS <sub>bg</sub>	(n <sub>b</sub> -1)( n <sub>g</sub> -1)	MS <sub>bg</sub> = SS <sub>bg</sub> / n <sub>bg</sub> -1	$\sigma_{(bg)}^2$
p x g	SS <sub>pg</sub>	(n <sub>p</sub> -1)( n <sub>g</sub> -1)	MS <sub>pg</sub> = SS <sub>pg</sub> / n <sub>pg</sub> -1	$\sigma_{(pg)}^2$
b x p x g, e	SS <sub>bpge</sub>	(n <sub>b</sub> -1)( n <sub>p</sub> -1)( n <sub>g</sub> -1)	MS <sub>bpge</sub> = SS <sub>bpge</sub> / n <sub>bpge</sub> -1	$\sigma_{(bpgp)}^2$

Gruplar arası farklılığın istatistiksel olarak anlamlı olup olmadığının test edilmesi için varyans bileşenlerinden F değerlerinin hesaplanması, G çalışmasında yapılmamaktadır (Shavelson ve Webb, 1991; Brennan, 2001).

Bu G-çalışmasına bağlı farklı farklı D-çalışmaları oluşturulabilir. Burada sadece, puanlayıcıların bir değişkenlik kaynağı olarak ele alındığı iki farklı D-çalışması açıklanmaya çalışılmıştır (Alharby, 2006):

- 1. D-Çalışması: İki görevin ve iki puanlayıcının bulunduğu, çaprazlanmış desen ve bağıl model düzenlenmiş olsun. Bu durumda bağıl varyans  $\sigma^2$ :

$$\sigma_{(\delta)}^2 = \frac{\sigma_{bg}^2}{N_g} + \frac{\sigma_{bp}^2}{N_p} + \frac{\sigma_{bpgp}^2}{N_g N_p} = \frac{\sigma_{bg}^2}{2} + \frac{\sigma_{bp}^2}{2} + \frac{\sigma_{bpgp}^2}{2 \times 2} \quad (2)$$

olacaktır. Burada  $\sigma_h^2$  gerçek varyanstır ve  $\sigma_{bg}^2, \sigma_{bp}^2$  ve  $\sigma_{bpgp}^2$  ise etkileşimsiz varyanslar olarak düşünülürler ve model bağıl olduğu için eşitlikte yer almazlar. Bu durumdaki G-

katsayısı, gerçek varyansın, gerçek varyans ile etkileşimli varyansların toplamına bölünmesiyle elde edilir:

$$G - \text{katsayısı} = \frac{\sigma_b^2}{\sigma_b^2 + \frac{\sigma_{bg}^2}{\lambda} + \frac{\sigma_{bp}^2}{\lambda} + \frac{\sigma_{bgp}^2}{\lambda x \lambda}} \quad (3)$$

Aynı durum için bağıl model yerine mutlak model kullanıldığı düşünüldüğünde, puanlayıcı, görev ve bunlar arasındaki etkileşime bağlı varyans bileşenleri mutlak varyansın parçaları olarak ele alınır ve önceki hesaplamada yer alan paydaya eklenerek daha büyük bir payda elde edilir. Bu durumda elde edilecek katsayının daha küçülmesine sebep olur. Böylece, güvenilirlik katsayısı denilen Phi katsayısı;

$$\begin{aligned} \Phi - \text{katsayısı} &= \frac{\sigma_b^2}{\sigma_b^2 + \frac{\sigma_g^2}{N_g} + \frac{\sigma_p^2}{N_p} + \frac{\sigma_{bg}^2}{N_g} + \frac{\sigma_{bp}^2}{N_p} + \frac{\sigma_{gp}^2}{N_g N_p} + \frac{\sigma_{bgp}^2}{N_g N_p}} \\ &= \frac{\sigma_b^2}{\sigma_b^2 + \frac{\sigma_g^2}{2} + \frac{\sigma_p^2}{2} + \frac{\sigma_{bg}^2}{2} + \frac{\sigma_{bp}^2}{2} + \frac{\sigma_{gp}^2}{2x2} + \frac{\sigma_{bgp}^2}{2x2}} \end{aligned} \quad (4)$$

şeklinde elde edilir.

(3) ve (4) eşitliklerinden puanlayıcıların sayısının artırılmasının ilgili hatayı düşüreceği ve sonuç olarak da güvenilirliği arttıracığı açıkça görülmektedir. Bunun anlamı, değerlendirmeye daha fazla sayıda puanlayıcı eklemek, genellenebilirlik kuramı çerçevesinde güvenilirlik katsayısını arttırmaktadır (Güler, 2008).

- 2. D-Çalışması: Aynı sayıda görev (2) ve puanlayıcıdan (2) oluşan, g x (p : b) olarak gösterilen bir karışık desen durumunu ele alalım.

Burada, sporcular görevlerle çaprazlanmış, ancak puanlayıcılar ile sporcular yuvalanmıştır. Bu demektir ki; her bir sporcu her bir görevi yerine getirirken, tüm görevleri gerçekleştiren farklı sporcuların performanslarını farklı puanlayıcılar

puanlamaktadır. Eğer modelin bağıl olmasına karar verilmişse, o zaman bağıl hata varyansı:

$$\sigma^2(\delta) = \frac{\sigma_p^2}{N_p} + \frac{\sigma_{bp}^2}{N_p} + \frac{\sigma_{bg}^2}{N_g} + \frac{\sigma_{gp}^2}{N_p N_g} + \frac{\sigma_{bgp}^2}{N_p N_g} = \frac{\sigma_p^2}{2} + \frac{\sigma_{bp}^2}{2} + \frac{\sigma_{bg}^2}{2} + \frac{\sigma_{gp}^2}{2x^2} + \frac{\sigma_{bgp}^2}{2x^2} \quad (5)$$

olacaktır. Bu durumda G-katsayısı;

$$G - \text{katsayısı} = \frac{\sigma_b^2}{\sigma_b^2 + \frac{\sigma_p^2}{\lambda} + \frac{\sigma_{bp}^2}{\lambda} + \frac{\sigma_{bg}^2}{\lambda} + \frac{\sigma_{gp}^2}{\lambda x \lambda} + \frac{\sigma_{bgp}^2}{\lambda x \lambda}} \quad (6)$$

olarak elde edilir.

Öte yandan, eğer modelin mutlak olmasına karar verilmişse, o zaman mutlak hata

$$\sigma^2(\Delta) = \frac{\sigma_p^2}{N_p} + \frac{\sigma_g^2}{N_g} + \frac{\sigma_{bp}^2}{N_p} + \frac{\sigma_{bg}^2}{N_g} + \frac{\sigma_{gp}^2}{N_p N_g} + \frac{\sigma_{bgp}^2}{N_p N_g} = \frac{\sigma_p^2}{2} + \frac{\sigma_g^2}{2} + \frac{\sigma_{bp}^2}{2} + \frac{\sigma_{bg}^2}{2} + \frac{\sigma_{gp}^2}{2x^2} + \frac{\sigma_{bgp}^2}{2x^2} \quad (7)$$

olacaktır.

Bu durumda güvenilirlik (Phi)katsayısı;

$$\Phi - \text{katsayısı} = \frac{\sigma_b^2}{\sigma_b^2 + \frac{\sigma_p^2}{\lambda} + \frac{\sigma_g^2}{\lambda} + \frac{\sigma_{bp}^2}{\lambda} + \frac{\sigma_{bg}^2}{\lambda} + \frac{\sigma_{gp}^2}{\lambda x \lambda} + \frac{\sigma_{bgp}^2}{\lambda x \lambda}} \quad (8)$$

olarak elde edilir.

Buradaki eşitliklerden de kolayca görülebileceği gibi, (6) eşitliğinin değeri (8) eşitliğinin değerinden daha büyüktür. Yine mutlak model için olan güvenilirlik katsayısı daha küçük olmaktadır. İlk durum ile ikinci durum karşılaştırılırsa, bağıl modelde, güvenilirlik çaprazlanmış desende daha büyükken, mutlak modelde güvenilirlik hem çaprazlanmış hem de yuvalanmış desen için aynı kalmaktadır (Güler, 2008).

Kısaca özetlenecek olursa, puanlayıcıların genellenebilirlik kuramı çerçevesinde (çaprazlanmış ya da yuvalanmış, bağıl ya da mutlak, tesadüfi ya da sabit) bir değişkenlik kaynağı olarak ele alınması, güvenilirlik katsayısının kestirilmesini etkilemektedir. Ayrıca, tek değişkenlik kaynaklı ve tümüyle çaprazlanmış desenler için hesaplanan genellenebilirlik katsayısı, puanlayıcılar arası güvenilirliğin tahmininde kullanılan (inter-class) sınıf içi korelasyon katsayısıyla aynı olmaktadır (Mushquash ve O'Connor, 2006).

### **1.5. PROBLEM DURUMU**

Sporun dünya üzerinde ulaştığı önem, bu öneme paralel olarak sporcular ve ülkeler arasındaki rekabetin boyutlarının her geçen gün biraz daha arttığı günümüzde, üstün yeteneğin mümkün olduğu kadar erken ve hatasız olarak bulunması sporda başarı için önem taşımaktadır.

Günümüzde sporda uluslararası platformlarda yarışmak ve dünyada bu alanda yerlerini almak için yetenekli sporculara ihtiyaç duyulmaktadır. Sporda potansiyel performansın erken yaşta belirlenmesi, sporcuların doğru spora yönlendirilmesi başarının elde edilmesine zemin hazırlayacaktır. Bunu sağlamak için de farklı branşlardaki performans kriterleri belirlenmeli, yetenek seçimi bu doğrultuda yapılmalıdır. Bunun sonucu olarak tüm disiplinlerde olduğu gibi spor alanında da bilimsel araştırmaların sayısı hızla artmaktadır. Öte yandan, yeteneği belirlemek için kullanılan testlerin yeterince güvenilir olmadığı gözlenmektedir. Yeteneğin sahada, salonda özel testlerle saptanmasının daha doğru olacağı, bunun ise yaş gruplarına uygun ve üzerinde özel olarak çalışılmış yöntemlerle geliştirileceği düşüncesi yaygındır.

Türkiye’de de özellikle son yıllardaki voleybol milli takımlarının başarılarından sonra voleybol, en çok ilgi duyulan spor dallarından biri haline gelmiştir. Bu yönelimin getirisi çocuklar ile gençleri zararlı alışkanlıklardan uzak tutmak, topluma kendine güvenen ve paylaşma duygusu gelişmiş üretken bireyler kazandırmak amacıyla spor kulüpleri bünyesinde alt yapı sporcuları yetiştirmektedir. Nitelikli voleybolcuları yetiştirmek amacıyla kulüplerin alt yapılarında voleybolcuların becerileri

gözlemlenmektedir. Becerilerin değerlendirilmesinde yoğun olarak gözlem yaklaşımı kullanılmaktadır. Bu bağlamda en uygun gözlemin yapılabilmesi amacıyla, puanlayıcıların (antrenörlerin), voleybolcuları puanlamaları ile elde edilen sonuçlarının tutarlılığının kuramsal olarak çalışılması, psikomotor becerilerin güvenilir biçimde ölçülmesi bakımından önemlidir.

Sporcuların, maçlar esnasında becerilerini en üst seviyede sergilemeleri gerekmektedir. Antrenmanlarla geliştirilmesi istenilen becerilerin voleybol sürecinin başında ve sürece devam ederken geçerli ve güvenilir ölçme araç ve yaklaşımlarının kullanılması ile ölçülmesi, voleybolcu hakkında verilecek kararlar bakımından önemlidir.

Performans değerlendirme görevlerinin belirli bir ölçeğe dayalı olarak birden fazla puanlayıcı tarafından puanlandığı ve elde edilen puanların da genellenebilirlik ve klasik test kuramına göre gerçek veriler üzerinde değerlendirildiği sınırlı ve avantajlı yönlerinin karşılaştırılması açısından bu çalışmanın önemli olduğu düşünülmektedir.

Bu araştırmada, performans değerlendirme görevlerinin bir gözlem formu ile birden fazla puanlayıcının puanlaması sonucu elde edilen puanlara klasik test kuramına dayalı ölçme yaklaşımının ve genellenebilirlik kuramının birlikte uygulanması ile hangisinin daha kullanılabilir olduğu, avantajlı olanının ve daha çok bilgi sağlayanının hangisi olduğunun belirlenmesi amacıyla aşağıdaki problem cümlesi ve alt problemlere cevap aranmaktadır.

## **1.6. PROBLEM CÜMLESİ**

Voleybol becerilerine ilişkin gözlem formunun birden fazla puanlayıcı tarafından puanlanması sonucu klasik test ve genellenebilirlik kuramı ile kestirilen parametreleri nasıldır?

## 1.7. ALT PROBLEMLER

### 1) Klasik Test Kuramına göre;

1.1.Voleybol Becerilerine İlişkin Gözlem Formuna ilişkin puanlamalar arası tutarlılığın derecesi nedir?

1.2.Voleybol Becerilerine İlişkin Gözlem Formuna ilişkin değerlendirmede dört farklı puanlayıcı arasındaki tutarlılık düzeyi nedir?

### 2) Genellenebilirlik Kuramına göre;

2.1.Voleybol Becerilerine İlişkin Gözlem Formu'ndan elde edilen puanların değerlendirmesinde kestirilen genellenebilirlik kuramı parametreleri nasıldır?

2.1.1. Tek değişkenli (performans değerlendirme tek boyuttan oluştuğu için) genellenebilirlik (G) çalışması sonucunda Voleybol Becerilerine İlişkin Gözlem Formu'ndan elde edilen puanların değerlendirmesinde kestirilen varyansları ve toplam varyansları açıklama yüzdeleri nasıldır?

2.1.2. Görev sayılarının 2, 8 ve 10, puanlayıcı sayılarının 1,2,3 ve 5 olması durumunda karar (K) çalışmasıyla kestirilen,

2.1.2.a. G katsayıları nasıl değişmektedir?

2.1.2.b. Phi katsayıları nasıl değişmektedir?

## 1.8. ARAŞTIRMANIN AMACI

Voleybolcuların becerilerinin tespit edilmesinde kullanılan Voleybol Becerilerine İlişkin Gözlem Formu'ndan elde edilen ölçmelerin; Klasik Test Kuramı ve Genellenebilirlik Kuramı'na dayalı olarak belirlenen güvenilirlik katsayılarının saptanması ve birbirleriyle karşılaştırılması araştırmanın amacını oluşturmaktadır.

## **1.9. ARAŞTIRMANIN ÖNEMİ**

Bu araştırmada, Voleybol Becerilerine İlişkin Gözlem Formu ile yapılan ölçmelerin güvenilirlikleri Klasik Test Kuramı ve Genellenebilirlik Kuramına dayalı olarak karşılaştırılmıştır.

Ağırlıklı olarak psikomotor becerilerin ölçmeye konu olduğu spor alanında, oyuncuların performanslarına ilişkin olarak yapılan geçerli çıkarsamaların olabildiğince hatadan arınık ve tutarlı olması da beklenir. Ancak bu koşullarda genç sporcuların uğraş verdikleri spor dallarında gelişimleri hakkında doğru yönlendirmeler yapıp, isabetli kararlar verilebilir, Spor eğitiminde ve spor etkinliklerinde ölçme konusu özelliklerle ilgili karara varma süreci, çoğu kez “hakem/antrenör” yargısı ile gerçekleşir. Bu durumda her ölçmede olduğu gibi, verilen kararların isabetliliğini etkileyecek temel unsurlar arasında ölçme aracının ve ölçütün niteliği dışında, puanlayıcı unsuru da öne çıkmaktadır. Bu bağlamda, puanlayıcıların psikomotor becerileri puanlamalarındaki güvenilirliğin test edilmesi önem kazanmaktadır. Spor bilimlerinde psikomotor davranışları ölçme ile ilgili literatürdeki çalışmalar doğru ve etkili değerlendirmeler üzerine yoğunlaşmaktadır. Araştırma bu yönüyle, psikomotor becerilerin ölçülmesinde uygulamacıların kuram tercihlerine bilimsel kaynaklık yapabilecektir.

Bunun yanı sıra, bu araştırma ile farklı branşlar için güvenilir dereceleme ölçeklerinin kullanılmasının önemini vurgulanarak, konunun spor bilimleri literatürü için yaygınlaşması da beklenmektedir.

## **1.10. SAYILTILAR**

Voleybolcular, pas, manşet, smaç, blok ve voleybol zekası voleybol becerilerini uygularken gerçek yeteneklerini yansıtmışlardır.

## **1.11. SINIRLILIKLAR**

Araştırma, voleybol becerilerinden olan pas, manşet, smaç, blok ve voleybol zekası becerilerini ölçmeyi amaçlayan 5 görevlik (25 maddelik ve 5 seçenekli) bir gözlem formundan elde edilen verilerle sınırlıdır.

## 1.12. KISALTMALAR

<b>G Kuramı</b>	: Genellenebilirlik Kuramı
<b>KTK</b>	: Klasik Test Kuramı
<b>G Çalışması</b>	: Genellenebilirlik Çalışması
<b>K Çalışması</b>	: Karar Çalışması
<b>G Katsayısı</b>	: Genellenebilirlik Katsayısı
<b>Phi Katsayısı</b>	: Güvenirlik Katsayısı
<b>b</b>	: Birey temel etkisi
<b>g</b>	: Görev (madde) temel etkisi
<b>p</b>	: Puanlayıcı temel etkisi
<b>bxg</b>	: birey ve görev ortak etkisi
<b>bxp</b>	: birey ve puanlayıcı ortak etkisi
<b>gxp</b>	: görev ve puanlayıcı ortak etkisi
<b>bxgxp</b>	: birey, görev ve puanlayıcı ortak etkisi

## 1.13. İLGİLİ ARAŞTIRMALAR

Bu çalışmada, KTK ve G Kuramı ile ilgili araştırmanın amacının belirlenmesine ışık tutan yurt içinde ve yurt dışında yapılmış çeşitli çalışmalara ve araştırmalara yer verilmiştir.

### 1.13.1.Yurt İçinde Yapılan Araştırmalar

Yelboğa'nın (2007) yaptığı araştırmada, 2005 ve 2006 yıllarında İş performansı ölçeği kullanılarak, klasik test kuramı ve genellenebilirlik kuramı analizleri gerçek verilerle uygulanmıştır. Klasik test kuramı ve genellenebilirlik kuramı analizleri sonucunda elde edilen verilerle, Klasik test kuramında ve Genellenebilirlik kuramında elde edilen güvenirlilik katsayıları karşılaştırılmıştır. Klasik test kuramında, test tekrar test ve Cronbach alfa güvenirlilik katsayıları, değerlendiriciler arası güvenirlilik için Kendall'ın konkordans katsayısı, genellenebilirlik kuramında ise çok değişkenli modelle G ve Phi katsayıları hesaplanmıştır. Elde edilen sonuçlara göre İş performansı ölçeğinde aynı

ölçme durumu için Klasik test kuramında ve Genellenebilirlik kuramının çok değişkenli modellerde elde edilen güvenilirlik katsayılarının birbirleriyle uyumlu sonuçlar ürettiği görülmüştür. Bu araştırma bulgularından yola çıkılarak elde edilen sonuçlara göre güvenilirliğin iş performansı ölçeğinde değerlendirilmesinde genellenebilirlik kuramının kullanılmasının uygun olacağı değerlendirilmiştir.

Güler'in (2008) yapmış olduğu araştırmada 2007 yılında yapılan matematik başarısının ölçülmesiyle elde edilen puanlara klasik test kuramı, genellenebilirlik kuramı ve çok değişkenli Rash ölçme modeli uygulanmıştır. Uygulanan bu üç kurama göre puanların güvenilirlikleri hesaplanmış ve üç kuramdan elde edilen sonuçlar karşılaştırılmıştır. Araştırmanın sonucunda, matematik başarısını ölçmek için kullanılan ölçme aracının güvenilir sonuçlar verdiği, puanlayıcıların birbirleriyle uyumlu puanlamalar yaptığı belirlenmiştir. Araştırma ile matematik başarısının ölçülmesinde güvenilirliğin belirlenmesinde yararlanılacak kuramların hangisinin seçileceği, elde edilen puanların hangi amaç için kullanılacağına bağlı olarak değişebileceği görülmüş, ancak araştırma sonuçlarına göre, matematik başarısının ölçülmesinde güvenilirliğin belirlenmesinde en az iki kuramdan yararlanmanın daha uygun olacağı sonucuna varılmıştır.

Atılgan (2008)'in çalışmasında, 2003-2004 eğitim-öğretim yılında İnönü Üniversitesi Eğitim Fakültesi Güzel Sanatlar Eğitimi Bölümü Müzik Öğretmenliği Programı özel yetenek seçme sınavlarının ikinci aşamasının güvenilirliği araştırılmıştır. Özel yetenek seçme sınavının birinci aşamada başarılı olarak ikinci aşama sınavına katılan 249 adaya üç alt boyutta uygulanan 13 görev dört puanlayıcı tarafından puanlanmıştır. Bu araştırmanın sonucu olarak bazı bireylerin bir maddeden diğerine bağlı durumlarının farklılaşmasının az olduğu, az da olsa bazı puanlayıcıların bazı bireyleri diğerlerine göre daha katı ya da cömert puanladıkları ve puanlayıcıların bireyleri bir görevden diğerine kararlı puanladıkları görülmektedir. Karar çalışmaları ile gözlem formu ve alt boyutların G ve Phi katsayıları hesaplanmıştır. Alternatif K çalışmaları ile G ve Phi katsayılarının artırılmak istenmesi durumunda, puanlayıcı sayısının artırılmasından çok madde sayılarının G ve Phi katsayılarını artırmada daha uygun olduğu sonucuna varılmıştır.

Nalbantoğlu'nun (2009) yaptığı araştırmada genellenebilirlik kuramına göre performans puanlanmada birden fazla puanlayıcı tarafından birlikte ve dönüşümlü olarak puanlanmasıyla oluşturulan desenlerden elde edilen G ve K çalışmaları karşılaştırılmıştır. Birlikte ve dönüşümlü olarak puanlanmasıyla oluşturulan  $\text{ö} \times \text{g} \times \text{p}$  ve  $(\text{ö}:\text{p}) \times \text{g}$  desenleri (ö:öğrenci, g:görev, p:puanlayıcı) için ayrı ayrı G ve K çalışmaları yapılmış ve bu değerler karşılaştırılmıştır. Araştırmada elde edilen sonuçlara göre,  $\text{ö} \times \text{g} \times \text{p}$  ve  $(\text{ö}:\text{p}) \times \text{g}$  desenleri ile yapılan analiz sonucunda her iki desende değişkenler için kestirilen varyans değerlerinin birbirleriyle paralel olduğu görülmüştür. Sınav sırasında öğrencileri puanlamada puanlayıcıların aralarında farklılıklar olmadığı bulunmuş ve her iki desene yapılan karar çalışmaları sonucunda, iki desendeki G ve Phi katsayıları arasında çok fark olmamakla birlikte  $(\text{ö}:\text{p}) \times \text{g}$  deseninde katsayıların daha büyük çıkma eğiliminde olduğu gözlenmiştir. Bu yüzden çok sayıda öğrencinin bulunduğu performans sınavlarında, puanlayıcılar arası tutarlık sağlandığında, öğrencilerin puanlayıcıların hepsi tarafından tek tek puanlanması yerine puanlayıcıların belli sayıdaki öğrencileri dönüşümlü olarak puanlamasının zaman, iş gücü ve ekonomiklik açısından daha uygun olacağı sonucuna varılmıştır.

2009 yılında Deliceoğlu'nun yaptığı araştırmada futbolcuların teknik yetilerinin tespit edilmesinde kullanılan Futbol Yetilerine İlişkin Dereceleme Ölçeği'nden elde edilen ölçmelerin; Klasik Test Kuramı ve Genellenebilirlik Kuramına dayalı olarak belirlenen güvenilirlik katsayılarının düzeyleri saptanmış ve birbirleriyle karşılaştırılmıştır. Çalışma grubunu, 2007-2008 yılları arasında Ankara ilindeki profesyonel spor kulüplerin futbol takımlarının alt yapılarındaki toplam 72 futbolcu oluşturmuştur. Futbol Yetilerine İlişkin Dereceleme Ölçeğinden elde edilen veriler ile ilk puanlama ve ikinci puanlama arasındaki tutarlılık için "Pearson momentler çarpımı" korelasyon katsayısı, maddelerin iç tutarlılık güvenilirliği için Cronbach  $\alpha$  (alfa) katsayısı, puanlayıcılar arasındaki tutarlılık için Kendall uyum katsayıları hesaplanmıştır. Ölçeğin alt boyutları ve ölçeğin toplamından elde edilen puanların güvenilirliği için puanlayıcı sayılarının 3, 4 ve 5 olduğu ve madde sayılarının bir arttırılıp, bir azaltıldığı koşullarda, G kuramında ana ve ortak etkilerin varyans bileşenlerin kestirilmesi için G ve Phi katsayıları bulunmuştur. Alt boyutlara ait sonuçlar incelendiğinde bütün alt boyutlar için en yüksek varyans bileşenin birey temel etkisi için kestirilen varyans bileşeni olduğu

görülmektedir. Bunu  $b_{mxp}$  ortak etkisi varyans bileşeni takip etmektedir. Ölçeğin puanlamasından elde edilen güvenirlik katsayıları iç ölçütlere göre incelendiğinde G katsayısı ve Cronbach Alfa katsayıları beklenen değerlerinden yüksek olduğu görülmekle beraber Phi katsayısı ile Kendall W Güvenirlik katsayılarının beklenen değerlerinden düşük olduğu görülmektedir. Bu bağlamda, araştırmadaki potansiyel hata kaynaklarının fazla olduğu durumlara benzer çalışmalarda Genellenebilirlik Kuramı, Klasik test kuramına alternatif oluşturmakta olduğu ve Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin güvenilir bir ölçek olarak değerlendirilebileceği sonucuna varılmıştır.

Güler'in (2009) yaptığı araştırmada genellenebilirlik kuramına ve terminolojisine ilişkin genel bir bakış açısı oluşturulmaya çalışılmıştır. Ayrıca, genellenebilirlik kuramına bağlı genellenebilirlik ve karar çalışmalarında elde edilen genellenebilirlik ve güvenirlik katsayılarının SPSS ile GENOVA paket programıyla elde edilen değerleri bir arada sunulmuştur. İki programa ilişkin bulunan sonuçların birbirine nerdeyse eşit değerler vermesi GENOVA'daki tüm bilgileri sağlayan, hatta bu bilgilere ek olarak mutlak ve bağıl hata kaynakları ile G ve phi katsayıları için ayrı ayrı grafiksel sonuçlar veren, kullanımı kolay SPSS tercih edilebilir bir alternatif olarak önerilebileceği sonucuna varılmıştır.

Taşdelen, Kelecioğlu ve Güler'in (2010) birlikte yapmış oldukları araştırmada Nedelsky ve Angoff standart belirleme yöntemleri ile elde edilen kesme puanlarının Genellenebilirlik Kuramı'na göre hesaplanan varyans bileşenleri karşılaştırılmıştır. Araştırmanın verileri 2008 yılı 6. sınıf Seviye Belirleme Sınavı'nda yer alan fen ve teknoloji sorularının Ankara ilinde görev yapan 40 fen ve teknoloji öğretmeni tarafından standart belirleme yöntemlerine uygun şekilde puanlanması ile elde edilmiştir. Kesme puanlarının G kuramı uygulamasıyla varyans bileşenleri ile G ve Phi katsayıları hesaplanmıştır. Araştırmanın sonunda, her iki yöntemle göre ayrı ayrı yapılan puanlamanın maddeler arası farklılığı ortaya çıkarabildiği belirlenmiştir. Angoff yöntemi ile belirlenen kesme puanları için puanlayıcılar arası tutarlılık görülmüş; Nedelsky yönteminde ise puanlayıcılar arasındaki uyumun düşük olduğu ortaya çıkmıştır. Nedelsky yöntemi ile elde edilen G ve Phi katsayıları daha yüksek olduğu

gözlenmiştir. Her iki standart belirleme yönteminin birlikte ele alındığı G çalışmasında ise maddeler arası farklılıkların ortaya çıkarılabildiği ve puanlayıcılar arasında bir tutarlılığın olduğu görülmüştür.

Yurt içinde yapılan sınırlı sayıda araştırmalara ilişkin olarak, performans değerlendirmede güvenilirliğin belirlenmesinde genellenebilirlik kuramı tercih edilebilir bir yaklaşım olarak görülmektedir. Performans değerlendirmede, hem birden fazla değişkenlik kaynağının birlikte göz önüne alınarak güvenilirliğin belirlenmesi hem de karar çalışmaları ile farklı sayıda değişkenlik kaynaklarına ilişkin önceden bilgi vermesi açısından genellenebilirlik kuramının kullanılması önerilmektedir.

### **1.13.2. Yurt Dışında Yapılan Araştırmalar**

Moon'un (1995) yaptığı çalışmada öğrencilerin performanslarına ilişkin nasıl güvenilir ölçmeler yapılabileceği ve kağıt-kalem testlerine karşı performans değerlendirmelerden nasıl farklı bilgiler elde edilebileceği açıklanmıştır. Bu çalışmada, üniversite öğrencilerine istatistik dersine ilişkin hem kağıt kalem testi ve hem de performans değerlendirme birlikte uygulanmıştır. Performans değerlendirmenin güvenilirliğinin kestirimi ve desenin geliştirilmesi için genellenebilirlik çalışması yapılmıştır. Puanların güvenilirliği hem klasik test kuramı ve hem de genellenebilirlik kuramı ile analiz edilmiştir. Her iki ölçme aracına ilişkin bağıl bilgiler elde etmek için korelasyonel ve açıklayıcı faktör analizi yapılmıştır. Araştırmanın sonucunda, performansları puanlayan iki puanlayıcının önemli bir hata kaynağı olmadığı görülmüştür. Ayrıca, öğrencilerin performanslarındaki tutarsızlığın başlıca kaynağının görevler olduğu gözlenmiş ve puanların güvenilirliğinin artırılabilmesi için görev sayısının artırılması gerektiği sonucuna ulaşılmıştır. İki puanlayıcı tarafından elde edilen tüm puanlar arasındaki ilişki ve G çalışmasının sonuçları, puanlayıcıların tutarlı bir şekilde öğrenci performanslarını puanlayabildiklerini göstermiştir. Çalışmanın sonucunda genellenebilirlik çalışmasındaki desende yer alan puanlayıcı değişkenlik kaynağının çıkarılması önerilmekle beraber bunun nedeni puanlayıcı sayısının bir azaltılabileceği olarak gösterilmektedir. İki ölçüm arasında oldukça yüksek korelasyon hesaplanmıştır. Faktör

analizi sonuçları faktör yapıları ile madde ayrıricılıkları arasında bir ilişki olduğunu göstermiştir.

Sub, Valiga ve Goa'nin (1997) birlikte yaptıkları çalışmalarında Akademik Danışmanlık Anketi (Academic Advising Survey, AAS)ne öğrencilerin verdiği puanların güvenilirliğini genellenebilirlik kuramına dayanarak araştırmışlardır. Bu çalışmada, lise sonrası eğitim veren 15 farklı enstitüde görev yapan 150 danışmanın her birinin 10 öğrenciye akademik danışmanlıkta bulunarak, danışmaya katılan 1500 öğrencinin 18 maddelik AAS'ne verdikleri puanlara (birey:akademi:danışman) $\times$ madde, rastgele etki modeline dayalı G ve K çalışması uygulanmıştır. G çalışması sonucunda en büyük değişkenlik kaynağının hata varyansı olduğu görülmüş ve bunu izleyen diğer en büyük iki değişkenlik kaynağı sırasıyla öğrenciler (birey) ve (akademi:danışman) değişkenlik kaynağı olarak bulunmuştur. K çalışması sonucunda, her bir danışmana düşen öğrenci sayısının 20'nin üzerine çıktığında güvenirlüğün oldukça yüksek olduğu gözlenirken, 10'dan düşük olduğunda AAS'nin güvenirlüğü düşük, 10 ile 20 arasında olduğunda güvenirlüğün orta düzeyde sonucuna varılmıştır. AAS'den elde edilen verilerin güvenirlüğü ayrıca klasik test kuramına dayalı olarak Cronbach alfa katsayısıyla hesaplanarak bulunmuştur. Ölçme sonuçları güvenirlüğünün araştırılmasında genellenebilirlik kuramının kullanılmasının olumlu ve olumsuz yönleri Cronbach alfa katsayısı çalışması ile karşılaştırılarak tartışılmıştır. Bu araştırmada yer alan anketin ve bu tür anket çalışmalarının güvenirlüğünün araştırılmasında, ölçme nesnesinin öğrenciler değil danışmanlar olduğu ve bu durumda danışmanların gerçek puan varyansının göz önüne alınması gerektiği vurgulanarak, Cronbach alfa katsayısı ile elde edilen güvenirlüğün yanlış yorumlanabileceği, bu durumda genellenebilirlik kuramına dayalı güvenirlilik çalışması yapılmasının daha uygun ve doğru olacağı yorumu yapılmıştır.

Kieffer'in (1998) yapmış olduğu çalışmasında genellenebilirlik kuramının niçin gerekli olduğunu, klasik test kuramının genelde yetersiz kaldığını; klasik test kuramına karşı genellenebilirlik kuramını kullanmanın yararlarını tartışmıştır. Çalışmada, genellenebilirlik kuramının klasik test kuramının bakış açısından yola çıkarak ve daha genişletilmiş bir hali olarak ortaya konulan bir kuram olduğu açıklanmıştır. Klasik test kuramında bir seferde sadece bir değişkenlik kaynağını (test tekrar-test, paralel formlar,

iç tutarlılık vb...) araştırmak mümkün olurken, aynı anda birden fazla değişkenlik kaynağı ve bunlar arasındaki etkileşimin hesaplanması mümkün olmazken, genellenebilirlik kuramının çoklu ölçme hatası kaynaklarının ve bunlar arası etkileşimin büyüklüklerini aynı anda tek bir analizle hesaplayabilmeye imkanı verdiğini açıklamıştır. Çalışmada klasik test kuramının sınırlılıkları ve genellenebilirlik kuramını kullanmanın güçlü yanları açıklanmış ve bu açıklamalar verilen hipotetik bir örnekte elde edilen sonuçlarla desteklenmiştir.

Tobar, Stegner ve Kane'nin (1999) birlikte yapmış oldukları çalışmada, 6 alt ölçek ve 65 maddeden oluşan ve kendi kendini değerlendirme ölçeği olan "Duygu Durum Profili" ölçeğinin güvenilirliğini G kuramı ile araştırmışlardır. Ölçeğinin 6 alt ölçeği için 0,74'den büyük G katsayıları elde edilirken toplam puanlardan elde edilen G katsayısı ölçeğin tesadüfi desenle sınanmasında 0,90 olarak bulunmuştur. Tesadüfi desende madde sayısı artırıldığında G katsayısının artma eğiliminde olduğu, bir ölçekteki madde sayısı ikiye katlandığında G katsayısının çok az arttığı ve alt ölçeklerin sabit değişken kaynağı olarak tanımlanması durumunda G katsayısının 0,96'ya yükseldiği sonucuna varılmıştır.

Lee ve Frisbie (1999) birlikte yapmış oldukları araştırmalarında maddeye dayalı güvenilirlik tahminlerinin test setlerinden (testlets) oluşan testler üzerinde yapılan önceki çalışmalarda abartılı sonuçlara ulaşmış olabileceği üzerinde durarak, bu tür bir testin güvenilirlik tahmininde genellenebilirlik teorisi yaklaşımının kullanılmasının gerekliliklerini ve uygunluğunu ortaya koymaya çalışmışlardır. Bu çalışmada, 1992 Bahar döneminde Temel Yetenek Iowa Testleri (Iowa Tests of Basic Skills) ve Eğitimsel Gelişim Iowa Testleri (Iowa Tests of Educational Development)'nin 4., 8. ve 11. sınıflara uygulanmasından elde edilen puanlar kullanılmıştır. Madde puanlarına dayalı Cronbach alfa kullanılarak elde edilen güvenilirlik tahminlerinin genellenebilirlik teorisi yaklaşımı kullanılarak elde edilene göre 0.04 kadar daha fazla olduğunu bulmuşlardır. Değişen pasaj sayılarından ve sabit toplam madde sayısına dayalı elde edilen genellenebilirlik katsayılarının, sabit pasaj sayılarından ve sabit toplam madde sayısına dayalı elde edilen genellenebilirlik katsayılarından daha değişken olduğu sonucuna varmışlardır. Dolayısıyla, etkili ölçme çalışmalarının gerçekleşmesi için

testlerdeki pasaj sayısını düzenlenmenin, her bir pasajdaki madde sayısını düzenlemekten daha verimli bir yol olacağını açıklamışlar, genellenebilirlik kuramının bu tür çalışmalarda sağladığı açıklayıcı bilgilerin önemini vurgulamışlardır.

Johnson, Penney ve Gordon (2000),birlikte yapmış oldukları çalışmada dereceli puanlama anahtarı (rubric) ile puanların çözümü ve puanlayıcılar arasındaki tutarlılığı araştırmışlardır. Araştırmada, puanlayıcıların uyumu; iki puanlayıcının puanlaması, orijinal puanlayıcıların puanları yerine bir uzmanın puanlarının kullanması ve iki puanlayıcı ile uzman puanlayıcının puanlarının kullanması alternatifleri sonucunda puanların kararlılığı incelenmiştir. Georgia Lisesi Yazma testinden alınmış 120 kompozisyon araştırmanın verileri olarak kullanılmıştır. İki puanlayıcının ve iki uzmanın içerik/organizasyon, tarz, yazım dili ve cümle oluşumu alanında verdikleri puanlarla puanlayıcılar arası güvenilirlik uyum yüzdesi, Spearman sıra farkları korelasyonu ve G kuramının Phi katsayısı ile hesaplanmıştır. Araştırma verilerine göre puanlayıcıların eğitilmesi ve puanlayıcı sayısının artırılmasının puanların güvenilirliğini artırdığı sonucuna varılmıştır.

Ludlow (2001) araştırmasında öğrencilerin okuma yetenekleri ile ilgilenmiş, değişkenlik kaynağı olarak öğrenci (s), pasaj (p) ve puanlayıcı (r) değişkenleri olarak değişkenlerin tümüyle çaprazlanmış ve her bir puanlayıcının farklı pasajları değerlendirdiği  $s \times (p:r)$  desenlerini kullanarak genellenebilirlik ve iki karar çalışması yapmıştır. Araştırmanın sonucunda elde edilen bulgulara göre güvenilir sonuçlara ulaşabilmek için iki puanlayıcı tarafından puanlanan her öğrencinin minimum üç pasaj okuması gerektiği tespit edilmiştir.

Goodwin (2001) araştırmasını hipotetik veriler kullanarak yapmıştır. Araştırmada basit yüzde ve kappa, basit korelasyon teknikleri ve Genellenebilirlik Kuramı yaklaşımlarını kullanmıştır. Çalışmasında, Pearson korelasyon katsayısı ile iki puanlayıcı arasında hesaplanan ve puanlayıcılardan birinin güvenilirliğini veren korelasyon katsayısı 0.90, basit yüzde ile hesaplanan puanlayıcıların farklı zamanlarda verdikleri puanların uyumu %10-70 aralığında (ortanca %20) ve puanlayıcıların uyumu %60-100 aralığında (ortanca %80) bulunurken, Kappa istatistikleri 0 ile 0.63 arasında hesaplanmıştır.

Genellenebilirlik kuramı ile hesaplanan G katsayısı ise 0.88 olarak bulunmuştur. Bu üç tekniğin hipotetik verilerle analiz edilmesi sonucunda kendi aralarındaki avantaj ve dezavantajlarına göre tartışılmış, bir çalışmada birçok hata kaynağını ayırabilmesi, daha kapsamlı ve esnek olması nedeniyle G kuramını araştırmalarda daha avantajlı olduğu vurgulanmıştır.

Kantor ve Lee (2007), Genellenebilirlik kuramını, TOEFL sınavında yazma becerilerinin ölçüldüğü bölüm için görev örnekleri ve alternatif değerlendirme yöntemleri belirlemek amacı ile kullanmışlardır. Araştırmada puanlayıcıların yazma becerilerine ait puanlardaki güvenilirliği ve farklı yazma görevleri araştırarak farklı senaryolara bağlı karar çalışması ile maksimum puan güvenilirliği için puanlayıcı ve görev sayısını belirlemeye çalışmışlardır. Çalışmanın sonuçlarına göre, maksimum puan güvenilirliği için görev başına düşen puanlayıcı sayısını arttırmaktansa sınavda kullanılan yazma görevlerinin sayısını arttırmanın daha etkili olduğu tespit edilmiştir. Ayrıca araştırmada okuma-yazma görevleri sayısının, dinleme-yazma görev sayısından daha az olduğunda daha yüksek güvenilirlik elde edileceği sonucuna varılmıştır.

Yurtdışında genellenebilirlik kuramına dayalı yapılan çalışmalar, değişkenlik kaynaklarının doğru ve açık bir şekilde tanımlandığı durumlarda genellenebilirlik kuramının açıklayıcı sonuçlar verdiğini göstermektedir. Performans değerlendirme çalışmalarında, genellenebilirlik kuramının klasik test kuramına göre daha kapsamlı ve aydınlatıcı bilgiler sunduğu, tercih edilebilir bir yöntem olduğu vurgulanmaktadır. Performans değerlendirmelerin genellenebilirlik kuramına dayalı güvenilirliğinin araştırılmasıyla, her bir değişkenlik kaynağına ve bunlar arasındaki etkileşime ilişkin değişkenliği hesaplayabilmenin, hem göreceli hem de mutlak güvenilirliği kestirebilmenin, farklı karar çalışmaları ile en uygun sayıda puanlayıcı, madde, ölçme durumuna karar vermenin mümkün olduğu farklı araştırmalarda elde edilen ortak sonuçlar arasında yer almaktadır.

## **BÖLÜM II**

### **YÖNTEM**

Bu bölümde, araştırmanın türü, verilerin toplandığı çalışma grubu, ölçme araçları, araştırma verileri ve verilerin analizine yer verilmektedir.

#### **2.1. ARAŞTIRMANIN TÜRÜ**

Araştırma, psikomotor becerilerin değerlendirilmesi çerçevesinde, puanlayıcıların tutarlılığının hesaplanmasında Klasik Test Kuramı ve Genellenebilirlik Kuramının avantaj ve dezavantajlarının ortaya konmasına, kuramlardan hangisinin daha çok bilgi verdiğinin saptanmasına ve iki kurama dayalı olarak elde edilen güvenilirlik katsayılarının karşılaştırmasına dayalı olup bu yönüyle karşılaştırmalı bir araştırmadır.

#### **2.2. ÇALIŞMA GRUBU**

Türkiye’de 13-14 yaş grubundaki lisanslı olarak voleybol oynayan erkek voleybolcular bu araştırmanın evrenini oluştururken, araştırmanın örneklem grubunu, 2009-2010 yılları arasında Türkiye Voleybol Federasyonu tarafından belirlenmiş olan 10 bölgede (Ankara- İstanbul- İzmir- Bursa- Çanakkale- Konya- Eskişehir- Antalya- Trabzon- Diyarbakır) görevlendirilen antrenörlerin seçmiş olduğu 102 erkek voleybolcu (ortalama yaş 13.8 yıl, antrenman yaşı 1 yıl ve üzeri) oluşturmuştur. Bu grubu oluşturan erkek voleybolcuların ailelerinden, kendilerinden ve takımlarının antrenörlerinden yeteneklerin ölçülmesi için önceden oluşturulmuş formlarla izinler alınmıştır. Voleybolcuların yetenek düzeyleri arasındaki farklılığı ortadan kaldırmak amacıyla çalışma grubunu antrenman yılları 1 ve üzeri olan voleybolcular oluşturmuştur.

#### **2.3. VERİ TOPLAMA ARACI**

Bu araştırmada; voleybolcuların becerilerinin (pas becerisi, manşet becerisi, smaç becerisi, blok becerisi ve voleybol (spor) zekası) ölçülmesi amacıyla alan uzmanları

tarafından 2007 yılında geliştirilen 25 maddeden oluşan “Voleybol Becerilerine İlişkin Gözlem Formu” kullanılmıştır.

Voleybol Becerilerine İlişkin Gözlem Formuna ait görevler için, voleybolda bazı becerileri kapsayan ve gözlem formunun alt boyutlarını oluşturan pas becerisi (k=5), manşet becerisi (k=5), smaç becerisi (k=5) , blok becerisi (k=5) ve voleybol (spor) zekası (k=5 ) becerilerine ait uzmanlar tarafından belirtilen kriterler (görevler) incelenmiştir.

- Pas Becerisi: Voleybol sporunda önemli bir teknik harekettir. Voleybolun en etkileyici ve en arzulanan hareketi olan smaç bile pasın başarısı ile değer kazanır. Pas, voleybol sporunda tamamen uzmanlık isteyen ve üzerinde çok çalışılması gereken bir harekettir. Bunun sebebi topun havaya atılması değil, isteğe göre belirli bir yere ve istenilen yüksekliğe atılmasıdır (Tiryaki, 1999).
- Manşet Becerisi: İki kolun iç tarafının yan yana gelerek, dirsekle el arasındaki bölümlerle topa vurma metodudur. Voleybol müsabakasında karşıdan gelen topların pasöre iletilmesi sırasında kullanılan bu hareket, başarılı yapılırsa servis karşılama ve savunma sırasında pasörü rahatlatır, takımın başarılı hücum yapma olanağını sağlar. Takımların oyun tempolarını manşetle ayarladıkları görülmektedir (Vurat,2000).
- Smaç Becerisi: Smaç sonuca ulaşabilmek için uygulanan en önemli teknik unsurlardan biridir. Genellikle smaç bir topu takip eden ve karşı takıma top geçmeden önce takımın topa son temasıdır. Smaç, file üzerinden karşı takımın sahasına doğru keskin bir açıyla patlayıcı bir şekilde topa sert vurularak yapılan dinamik bir harekettir.Smaç takımın sayı almak için ve karşı takımın üzerinde güç kazanmak için hücumda en öldürücü silahtır (İskender, 2002).
- Blok Becerisi: Blok zamanlama ve rakip smaçörün yaklaşımlarını okuyabilme yeteneğini içerir. Bloкта amaç sert vurulan topu önlemek ve karşı alana geri göndermek veya smacın hızını yavaşlatmak ya da kesmektir (Fröhner, 1999).

- Voleybol (spor) Zekası: Voleybol teknik ve süratli bir oyun olup doğru ve hızlı karar verme yeteneği gelişmiş sporcuları bünyesinde barındırır. Voleybolcular, oyunu okuyup karşı takımın yapacağı kombinasyonları saliseler içinde fark edip buna karşı önlem almak zorundadırlar ( Tiryaki, 1999).

“Voleybol Becerilerine İlişkin Gözlem Formu” voleybol branşındaki iki akademisyen ve ölçme ve değerlendirme konusunda bir uzman tarafından incelenmiş ve formun maddeleri önerilere göre hazırlanmıştır. Bu inceleme sırasında gözlem formunun madde sayıları değişmemiştir. Böylece gözlem formunun, ölçülmek istenilen voleybol becerilerini kapsayıp kapsamadığı uzman görüşlerine dayalı olarak belirlenmiştir.

Performansa ait becerilerin hiyerarşik bir sırayla yapılmasının gerekliliğinden dolayı Voleybol Becerilerine İlişkin Gözlem Formuna ait maddelerin sıraları değişmemiştir. Gözlem Formundan alınabilecek en düşük puan 0 ve en yüksek puan 100 arasında değişmektedir. Voleybol Becerilerine İlişkin Gözlem Formunun alt boyutlarından alınabilecek en düşük ve en yüksek puanlar incelendiğinde pas becerisi 0-20, manşet becerisi 0-20, smaç becerisi 0-20 , blok becerisi 0-20 ve voleybol (spor) zekası 0-20 arasında değişmektedir. Voleybol Becerilerine İlişkin Gözlem formuna ait maddelerde belirtilen becerilerden, beş kategorili formda puanlayıcılardan “Görülmedi (Çok kötü)”, “Yetersiz (Kötü)”, “Yeterli (Orta)”, “Oldukça Yeterli (İyi)” ve “Mükemmel (Çok iyi)” seçeneklerinden birini işaretlemeleri istenmiştir. Ölçülmek istenilen beceriyi tam anlamıyla gerçekleştiren voleybolcunun alacağı “Mükemmel (Çok iyi)” derecelemesi “4” ve voleybolcu tarafından becerinin gerçekleştirilmemesi durumunda alacağı “Görülmedi (Çok kötü)” ibaresine karşılık gelen “0” arasında değişen puanlar belirlenmiştir. Ölçeğin bütünü ve alt boyutlarına ait maddelerden alınan puanların yükselmesi, voleybol becerisininin yüksek olduğunu ifade etmektedir.

#### **2.4. VERİLERİN TOPLANMASI**

Araştırmanın amacına uygun olarak becerilere ait ölçmeler çalışma grubuna katılan Türkiye’deki 13-14 yaş grubundaki lisanslı olarak voleybol oynayan erkek voleybolcular önceden belirtilen fikstür doğrultusunda maçlar oynamışlardır. Maçların

oyunması sırasında video-kamera ile elde edilen görüntüler, becerilerdeki bütün hareketlerin rahat görünebileceği şekilde daha önceden video kamera çekimi konusunda eğitimi verilmiş uzman kişiler tarafından salonların iki farklı yönünden her maç çift kamera ile çekilmiştir.

Gözlem formundaki beceriler 4 puanlayıcı tarafından incelenmiş ve görevlerin puanlanması sırasında dikkat edilmesi gereken kriterler belirlenmiştir. Puanlama yapılmadan önce tanıtım amaçlı uzman bir kişi tarafından üç kez sunum yapılmış ve puanlamada nelere dikkat etmeleri gerektiği vurgulanmıştır. Her puanlayıcı sporcuların (102 sporcu) hepsini puanlamıştır. Puanlama çıplak gözle izlenen maçlar sırasında, puanlayıcıların birbirlerinden etkilenmemelerini sağlamak amacı ile kendileri için özel olarak birbirlerinden uzak hazırlanmış yerlerde maçları takip etmişlerdir. Her puanlayıcı sporcuların üç farklı maçını izleyerek puanlama yapmışlardır. Elde edilen görüntüler antrenörlük yılları 10 ve üzeri olan 4. kademe voleybol antrenörlük belgesine sahip 4 uzman tarafından değerlendirilmiştir. Puanlayıcılar istedikleri taktirde elde edilen video kamera çekimlerinden maçları tekrar seyretmişlerdir.

## **2.5. VERİLERİN ANALİZİ**

Voleybol Becerilerine İlişkin Gözlem Formunun performans değerlendirmesini yapabilecek dört voleybol uzmanı görev almıştır. Bu puanlayıcılar tamamen birbirlerinden bağımsız puanlama yapmışlardır. Verilerin analizi iki aşamada gerçekleştirilmiştir. İlk aşamada, dört puanlayıcıya göre ayrı ayrı, 102 sporcu için 5 göreve verilen puanların betimsel istatistikleri hesaplanmıştır. İkinci aşamada ise veriler araştırmanın görevlerine göre klasik test kuramı ve genellenebilirlik kuramı için kullanılan desene ( $b \times g \times p$ ) uygun olarak analiz edilmiştir.

Voleybol Becerilerine İlişkin Gözlem Formuna ait maddelerin kendi içinde tutarlı ölçme yapıp yapmadığının belirlenmesi amacıyla maddelerin iç tutarlılık güvenilirliği için Cronbach  $\alpha$  (alfa) katsayısı hesaplanmıştır. Puanlayıcıların Voleybol Becerilerine

İlişkin Gözlem Formuna puanlanmaları arasındaki tutarlılık için Kendall'ın konkordans (uyuşum) katsayısı hesaplanmıştır.

Voleybol Becerilerine İlişkin Gözlem Formunun güvenilirliği, puanlayıcı sayılarının 1,2, 3, 4 ve 5 olduğu ve görev sayılarının arttırılıp 8 ve 10, azaltılıp 2 olduğu koşullarda, G kuramında temel ve ortak etkilerin varyans bileşenlerinin kestirilmesi için G ve Phi katsayıları bulunmuştur. Elde edilen varyans eşitlikleri aşağıdaki belirtilen kriterlere uygun olarak yorumlanmıştır.

Birey (b) temel etkisi için kestirilen varyans bileşenin büyük olması, bireylerin ölçülen özellikleri bakımından ayrılabilirliklerini göstermektedir. Görev(g) temel etkisi için kestirilen varyans bileşenin yüksek olması bazı görevlerin diğerlerine göre güçlük düzeylerinin daha yüksek olduğunu, sıfıra yakın olması görev güçlük düzeylerinin birbirine yakın olduğunu göstermektedir. Puanlayıcı (p) temel etkisi için varyans bileşenin yüksek olması bazı puanlayıcıların diğerlerine göre farklı puanlama yapması, sıfıra yakın olması ise puanlayıcıların birbirine benzer puanlama yaptığını göstermektedir (Shavelson ve Webb,1991).

Birey-görev ortak etkisi (bg); bireyin bağıl konumunun bir görevden diğerine değişip değişmediğini bir başka deyişle, belli bir görevin, tüm bireyler için aynı olup olmadığı ya da belli bir birey için anlamlı olarak daha zor veya daha kolay olup olmadığını gösterir. Varyans bileşenin büyük olması bireylerin bir görevden diğer göreve bağıl konumlarının değiştiği şeklinde yorumlanabilir. Birey-puanlayıcı ortak etkisi (bp) belli bir puanlayıcının, tüm bireyler için aynı şekilde puanlama yapıp yapmadığı veya belli bir bireyin puanlanmasıyla diğer bireylerin puanlanması bakımından bağımlılık olup olmadığını belirtir. Birey-puanlayıcı ortak etkisi varyans bileşenin sıfır ya da sıfıra yakın olması hiçbir puanlayıcının bireyler arasındaki puanlamasında birbirine bağımlı olmadığını göstermektedir. Görev-puanlayıcı ortak etkisi (gp) belli bir puanlayıcının tüm görevler için aynı biçimde veya belli bir görev için diğerlerinden farklı davranıp davranmadığını, puanlayıcıların puanlamalarının bir görevden diğerine kararlı olup olmadığını ya da etkilenip etkilenmediğini göstermektedir. Kestirilen varyans

bileşenin sıfır ya da sıfıra yakın olması durumunda puanlayıcıların bir görevden diğerine puanlamalarında kararlı olduklarını göstermektedir (Shavelson ve Webb,1991).

Kalan etki varyansı (bgp) Birey-görev-puanlayıcı ortak etkisi ve ölçülemeyen varyans kaynağı, tesadüfi hatayı göstermektedir. Varyans bileşen değerinin sıfıra yakın olması b<sub>x</sub>g<sub>x</sub>p birleşiminden kaynaklanan hata kaynağının ya da ölçülemeyen hata kaynaklarının az olduğu, varyans bileşenin büyük olması ölçülemeyen hata kaynağının büyük olduğu şeklinde yorumlanabilir (Shavelson ve Webb,1991).

## BÖLÜM III

### BULGULAR VE YORUM

Bu bölümde, araştırmanın problem cümlesi ve alt problemlerin sırasına göre elde edilen bulgular ve bulgulara ilişkin yapılan yorumlar yer almıştır.

#### 3.1. Alt Problem I: Klasik Test Kuramına Göre Voleybol Becerileri Performansının Değerlendirilmesinin İncelenmesi

##### 3.1.1. Klasik Test Kuramına Göre Voleybol Becerileri Performansının Değerlendirilmesinin İç Tutarlılığı

Sonuçların daha iyi anlaşılabilmesi için Tablo:2’de sporcu puanlarının dört puanlayıcıya ilişkin betimsel istatistiklerine yer verilmiştir.

**Tablo:2 Sporcu Puanlarının Dört Puanlayıcıya İlişkin Betimsel İstatistikleri**

İstatistikler	Puanlayıcılar			
	1	2	3	4
N	102	102	102	102
Ortalama	27.80	13.99	15.01	38.17
Std. Sapma	12,65	5.50	5.50	13.44

Tablo 2’de görüldüğü gibi, 102 sporcunun 5 görevden (25 madde) 0 - 4 puan ölçeği üzerinden aldıkları puanlara ilişkin en yüksek ortalama 38.17 ile 4. puanlayıcıya ve en düşük ortalama 13.99 ile 2. puanlayıcıya aittir. Diğer iki puanlayıcıya ait ortalama değerler 15.01 ve 27.80 dir. Standart sapmalara bakılacak olursa 2. ve 3. puanlayıcıların standart sapmaları aynı ve 5.50 olarak hesaplandığı görülmektedir. En yüksek standart sapma 4. Puanlayıcıya ait olup 13.44’ tür. 1. Puanlayıcının standart sapması ise 12.65

olarak hesaplanmıştır. Elde edilen bu verilere göre 4. puanlayıcı daha cömert, 2 ve 3. puanlayıcıların daha katı oldukları yorumu yapılabilir.

**Tablo:3 Sporcu Puanlarının Dört Puanlayıcıya İlişkin İç Tutarlılığı**

	Puanlayıcılar			
	1	2	3	4
Cronbach $\alpha$	0.85	0.88	0.89	0.92

Tablo 3’de görüldüğü gibi sporcuların “Voleybol Becerilerine İlişkin Gözlem formu” ile değerlendirilmesinden aldıkları puanların dört puanlayıcıya göre güvenilirlik analizleri yapılmıştır. Performans değerlendirmelerin yapısına ilişkin temel varsayımlardan biri de değerlendirmede yer alan her bir maddenin gözlenmek istenen performansla ilişkili olmasıdır. Bu her bir maddenin istenilen performansı ölçebildiğini ifade eder. Bu sebeple, performans değerlendirmede yer alan maddelerin iç tutarlılıklarının belirlenmesi gerekmektedir. Bu amaca yönelik olarak da her bir puanlayıcının verdiği puanlar için Cronbach  $\alpha$  güvenilirlik katsayısı hesaplanmıştır. Birbiriyle yüksek ilişki gösteren maddelerin  $\alpha$  güvenilirlik katsayısının da yüksek olması beklenmiştir. Voleybol becerileri performansının değerlendirilmesine ilişkin Cronbach  $\alpha$  güvenilirlik katsayıları birinci puanlayıcı için 0.85, ikinci puanlayıcı için 0.88, üçüncü puanlayıcı için 0.89 ve dördüncü puanlayıcılar için 0.92 olarak bulunmuştur. Elde edilen bu katsayılara göre, voleybol becerileri performansını değerlendirmede yer alan maddelerin voleybol becerilerini birbirleriyle puanlayıcıların verdiği puanlar kendi içinde tutarlıdır yorumu yapılabilir.

### **3.1.2. Klasik Test Kuramına Göre Voleybol Becerileri Performansının Değerlendirilmesinde Dört Farklı Puanlayıcının Puanları Arasındaki Tutarlılık Düzeyi**

Voleybol becerileri performansının değerlendirilmesinde, dört farklı puanlayıcının aynı koşullar altında her bir sporcuyla puanlamasına ilişkin elde edilen puanlar arasındaki

tutarlılık derecesi, parametrik olmayan istatistiksel bir teknik olan Kendall'ın Uyuşum Katsayısı (Kendall's Concordance Coefficient) ile analiz edilmiş ve her bir puanlayıcının verdiği puanlar ile diğer bir puanlayıcının verdiği puanlar arasındaki korelasyon katsayıları hesaplanmıştır. Puanlayıcıların 5 görev üzerinden verdikleri puanlar arasındaki korelasyon değerleri Tablo 4'te verilmiştir.

**Tablo:4 Dört Puanlayıcının 5 Göreve Verdikleri Puanlar Arasındaki Korelasyon Katsayıları**

	1.Puanlayıcı	2.Puanlayıcı	3.Puanlayıcı	4.Puanlayıcı
1.Puanlayıcı	-	0.335	0.283	0.719
2.Puanlayıcı		-	0.443	0.218
3.Puanlayıcı			-	0.177
4.Puanlayıcı				-

Tablo 4'te verilen sonuçlara göre puanlayıcılar arası uyumun oldukça düşük olduğu gözlenmiştir. Ayrıca, 5 görev (25 madde) için Kendall'ın Uyuşum Katsayısı (Kendall's Concordance Coefficient) 0.37 olarak bulunmuştur ( $X^2=224.25$  sd=3,  $p=0.00<0.05$ ). Elde edilen bu sonuca göre puanlayıcılar arası uyum olmadığı yorumu yapılabilir (Howell, 2002). Puanlamalar yapıldıktan sonra puanlayıcılar ile görüşme yapılmıştır. Bütün puanlayıcılara puanlama yapılmadan önce üç kere puanlamanın nasıl yapılması gerektiğine dair eğitim verilmiş olmasına rağmen, puanlayıcılar geçmiş sporculuk yaşantılarında oynadıkları mevkiye göre uzmanlaştıklarını ve puanlamalarına bunu yansıtıklarını ifade etmişlerdir. Örneğin sporculuğunda pasör olarak oynamış olan 1. ve 4. Puanlayıcının her ikisi de değerlendirmelerini, bu bakış açısıyla yaptıklarından dolayı birbirleriyle tutarlılıkları en yüksek değer olan 0.719 değerini elde etmişlerdir. Bu puanlayıcılar pas görevinin puanlanmasında istenilen "Topu başının üstünde yakalıya biliyor mu?" sorusunun karşılığında bazı sporcuların bu beceriyi gösterdiklerini fakat

buna rağmen parmak kuvvetleri yetmediğinden topu istenilen hedefe yönlendiremediklerini ve bunun sonucunda düşük puanla değerlendirildiklerini ifade etmişlerdir. Diğer iki puanlayıcıdan birinin sporculuk geçmişinde libero, diğerinin smaçör olması ve kendi uzmanlıklarının puanlara yansımaları puanlayıcıların ifadelerinde yer almaktadır. Bu ifadelerin sonucunda puanlayıcılar arası tutarlılığın düşük çıkmasının nedenlerinden birinin puanlayıcıların uzmanlık alanına göre puanlama yapmış oldukları şeklinde yorumlanmıştır. Uzmanlık alanlarının farklı olması görevlere farklı şekilde algılamalarına neden olmaktadır.

Woods (1997) performans ölçümünde yapılan hataları sekiz başlık altında incelemektedir. Bu çalışma bulgularının da yukarıda belirtilen hata kaynaklarından etkilendiği düşünülmeyle birlikte, puanlayıcıların, bilgi düzeylerindeki ve sporcularla olan geçmişlerindeki farklılıktan dolayı, taraflı ölçüm ya da katılık / cömertlik ten kaynaklanan hatalardan daha çok etkilenmesine neden olabileceği düşünülmüştür.

### **3.2. Alt Problem 2: Genellenebilirlik Kuramına Göre Voleybol Becerileri Performansının Değerlendirilmesinin İncelenmesi**

#### **3.2.1.Voleybol Becerileri Performansının Değerlendirilmesinin Kestirilen Genellenebilirlik Kuramı Parametreleri**

Voleybol Becerileri performansının değerlendirilmesinin G çalışması ile elde edilen varyanslarını ve varyans yüzdelerini hesaplamak amacıyla, 2007 yılında geliştirilmiş olan “Voleybol Becerilerine İlişkin Gözlem formu”ndeki 5 göreve (25 madde) tümüyle çaprazlanmış  $b \times g \times p$  modeli uygulanmıştır. Değerlendirmenin uygulandığı 102 sporcu (birey), 4 puanlayıcı ve 5 görevden oluşan orjinal verilerle tek değişkenli modelle yapılan G çalışması için; kestirilen varyans bileşenleri ve toplam varyansı açıklama yüzdeleri  $b$ ,  $g$  ve  $p$  ana etkileri ile  $bg$ ,  $bp$ ,  $gp$  ve  $bgp$  ortak etkileri Tablo 5’te verilmiştir.

**Tablo:5 Voleybol Becerileri Performansının Değerlendirilmesinin G Çalışması Sonucunda Değerlendirmenin Kestirilen Varyansları ve Toplam Varyansı Açıklama Oranları**

Varyans Kaynağı	Sd	Toplam Kareler	Kareler Ortalaması	Varyans	%
b	101	9503.473	94.094	2.603	2.5
g	4	40021.627	10005.407	0.679	0.5
p	3	45514.327	15171.442	8.982	5
bg	404	18049.373	44.677	0.452	0.3
bp	303	15184.423	50.114	0.827	0.7
gp	12	135031.435	11252.620	103.339	62
bgp	1212	64544.565	53.255	50.122	29
<b>Toplam</b>					<b>100</b>

Tablo 5'te verilen tek değişkenli G çalışması sonucunda voleybol becerileri performansını değerlendirmenin kestirilen varyans ve toplam varyansı açıklama oranları incelendiğinde, birey (b) ana etkisi için kestirilen varyans bileşeninin (2.603) toplam varyansın % 2.5'ini açıkladığı görülmektedir. Burada bireyler için kestirilen varyans bileşeni, toplam varyans içinde dördüncü sırada paya sahip varyans bileşenidir. Genellenebilirlik çalışmalarında, birey ana etkisi evren puanı varyansı olarak değerlendirilir ve ölçülen özellik açısından bireyler arası farklılaşmayı ifade eder (Shavelson ve Webb, 1991; Brennan 2001). Bu sonuca göre bireylerin birbirinden ayrılamadığı yorumu yapılabilir.

Bireyler için kestirilen varyans bireyler arası farklılıklar olarak alınır. Bireyler için kestirilen varyansın toplam varyans içindeki oranının en büyük olması istenilen bir durumdur. Ancak burada istenilen durum gerçekleşmemiş ve birey (b) ana etkisi için tek değişkenli modelle yapılan G çalışmasından kestirilen varyans bileşeni (2.603) toplam

varyansın % 2.5'ini açıklamıştır. Birey ana etkisinin varyans bileşeni büyüklüğünün, diğer ana etkiler içinde ikinci sırada yer aldığı görülmektedir. Bu durum, ölçme ile elde edilen boyutta bireyler arası farklılıkların ortaya çıkarılmadığının bir göstergesi olmuştur. Görev(g) temel etkisi için kestirilen varyans bileşenin yüksek olması bazı görevlerin diğerlerine göre güçlük düzeylerinin daha yüksek olduğunu, sıfıra yakın olması görev güçlük düzeylerinin birbirine yakın olduğunu göstermektedir. Görev (g) ana etkisi için tek değişkenli modelle yapılan G çalışmasından kestirilen varyans bileşeni (0.679) toplam varyansın % 0.5'ini açıklamaktadır. Görev ana etkisinin varyans bileşeni büyüklüğünün, diğer ana etkiler içinde son sırada yer aldığı görülmektedir. Bu durum görev güçlük düzeylerinin birbirine yakın olduğu şeklinde yorumlanabilir. Puanlayıcı ana etkisi için kestirilen varyans bileşeni (8.982) toplam varyansın % 5'ini açıklayarak, ana etkiler içinde en yüksek varyans bileşeni olmaktadır. Puanlayıcı ana etkisinin G çalışması ile kestirilen varyans oranının yüksek çıkması, puanlayıcıların tüm bireyler için yaptıkları puanlamalar arasında çokça bir farklılık bulunduğunun, puanlamalar arasında bir tutarlılığın söz konusu olmadığını göstergesi olabilir.

Birey x görev (bg) ortak etkisi (0.452) toplam varyansın % 0.3'ünü açıklamaktadır. Birey x görev ortak etkisi tek değişkenli analizle elde edilen en küçük varyanstır. Bu durum, tek değişkenli analizle bu ölçme için birey x görev ortak etkisinden kaynaklanan farklılığın olmadığını, belli bireylerin bağlı durumlarının bir görevden diğerine çok farklılaşmadığını göstermektedir. Birey x puanlayıcı (bp) ortak etkisi (0.827) toplam varyansın % 0.7'sini açıklamaktadır. Birey x puanlayıcı ortak etkisi tek değişkenli analizle elde edilen en küçük varyanslardan biridir. Buna göre, tek değişkenli analizle bu ölçme için birey x puanlayıcı ortak etkisinden kaynaklanan farklılığın olmadığı, belli bireylerin bağlı durumlarının bir puanlayıcıdan diğerine çok değişmediği yorumu yapılabilir. Başka bir ifadeyle, birey x puanlayıcı ortak etkisi varyans bileşenin sıfır ya da sıfıra yakın olması hiçbir puanlayıcının bireyler arasındaki puanlamasında birbirine bağımlı olmadığını göstermektedir yorumu yapılabilir. Görev x puanlayıcı (gp) ortak etkisi (103.339) toplam varyansın % 62'sini açıklamaktadır. Görev x puanlayıcı ortak etkisi tek değişkenli analizle elde edilen en büyük varyanstır. Bu sonuca göre, tek değişkenli analizle bu ölçme için görev x puanlayıcı ortak etkisinden kaynaklanan farklılığın çok büyük olduğu, belli görevlere verilen puanların puanlayıcıdan

puanlayıcıya çok fazla değiştiği görülmektedir. Bu değişimin puanlayıcıların uzmanlık alanlarının birbirinden farklı olması ve her puanlayıcının uzmanlık alanlarına göre değerlendirme yaptıkları şeklinde yorumlanabilir. Ayrıca puanlayıcıların gözlem formunu kullanma becerilerinin yetersiz olması ve gözlem konusundaki eksiklikleri puanlayıcıdan puanlayıcıya puanların birbirinden farklı olmasının nedeni olabilir. Birey x görev x puanlayıcı (artık) ortak etkisi varyans bileşeninin de (50.122) toplam varyansı açıklama oranı % 29 olarak görülmektedir. Bu oran tablodaki ikinci en büyük değerdir. Birey x görev x puanlayıcı (artık) varyansın büyük olması; birey, puanlayıcı ve görev ortak etkisi ve /veya tesadüfi hataların büyük olabileceğinin bir göstergesi olabilir.

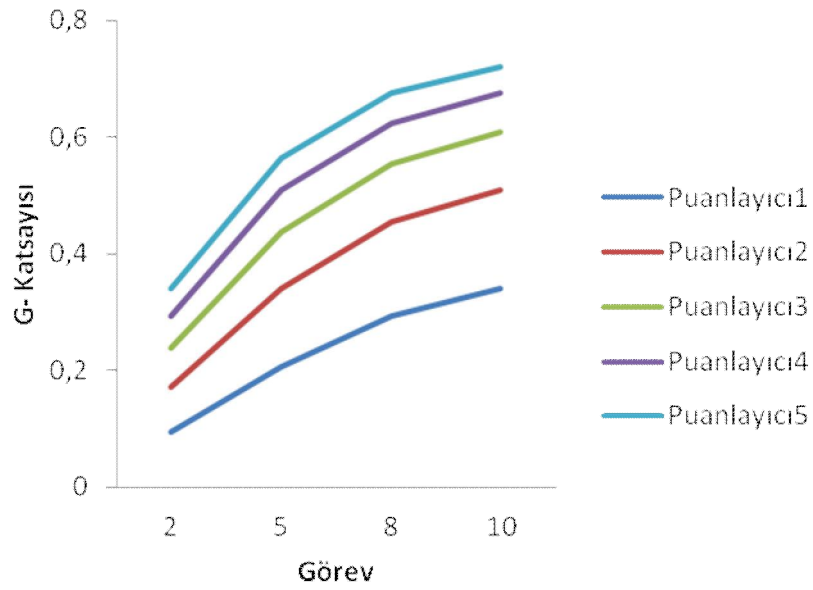
### 3.2.2. Görev sayılarının 2, 5, 8 ve 10, puanlayıcı sayılarının 1, 2, 3, 4 ve 5 olması durumunda karar (K) çalışmasıyla kestirilen, G ve Phi Katsayıları

“Voleybol Becerilerine İlişkin Gözlem Formundaki 5 görev (25 madde) sayısına ve artırılıp azaltılarak 2, 5, 8 ve 10 olduğu, puanlayıcı sayısına ve artırılıp azaltılarak 1, 2, 3, 4 ve 5 olduğuna göre düzenlenen senaryolar için yapılan K çalışması sonucunda kestirilen G ve Phi katsayıları varyansları Tablo 6, Şekil 1 ve Şekil 2’de verilerek açıklanmıştır.

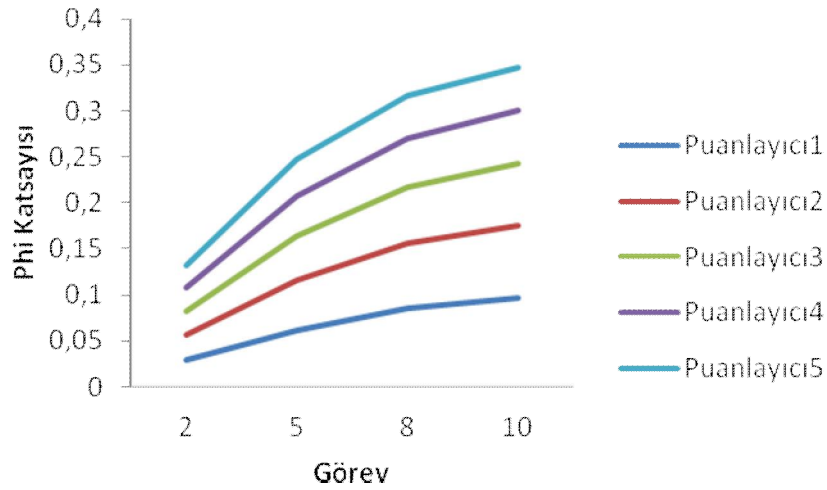
**Tablo:6 Voleybol Becerileri Performansının Değerlendirilmesinin K Çalışması ile Madde ve Puanlayıcı Sayıları Senaryolarına Göre G ve Phi Katsayıları**

Görev Sayıları	Puanlayıcı Sayıları									
	1		2		3		4		5	
	G	$\Phi$	G	$\Phi$	G	$\Phi$	G	$\Phi$	G	$\Phi$
2	0.094	0.029	0.172	0.057	0.238	0.083	0.293	0.108	0.342	0.132
5	0.206	0.062	0.342	0.116	0.438	0.164	<b>0.509</b>	<b>0.208</b>	0.565	0.247
8	0.293	0.085	0.454	0.156	0.555	0.217	0.624	0.270	0.675	0.316
10	0.342	0.097	0.509	0.176	0.609	0.243	0.675	0.300	0.722	0.348

Tablo 5, Şekil 1 ve Şekil 2’de görüldüğü gibi, voleybol becerileri performansını değerlendirmenin 5 görev (25 madde) sayısına ve bu maddeleri puanlayan 4 puanlayıcıya göre elde edilen G katsayısı 0.509 ve  $\Phi$  katsayısı da 0.208 olarak kestirilmiştir. Buradan görüleceği gibi, uygulamadaki durumda değerlendirmeden elde edilen puanların G katsayısı  $\Phi$  katsayısından daha yüksektir. Görev sayısı sabit tutularak (5) puanlayıcı sayısının bir azaltılması (3) durumunda kestirilen G katsayısı 0.438 ve  $\Phi$  katsayısı 0.164, puanlayıcı sayısının bir arttırılması (5) durumunda ise G katsayısı 0.565 ve  $\Phi$  katsayısı 0.247 olarak kestirilmiştir. Puanlayıcı sayısı sabit tutularak (4) görev sayısının ikiye düşürülmesi durumunda kestirilen G katsayısı 0.293 ve  $\Phi$  katsayısı 0.108, görev sayısının iki katına çıkarılması (10) durumunda ise G katsayısı 0.675 ve  $\Phi$  katsayısı 0.300 olarak hesaplanmıştır. Buradan da görüleceği üzere, puanlayıcı sayısı sabit tutularak görev sayısının iki katına çıkması durumunda G katsayısındaki artma 0.166 iken, görev sayısı sabit tutularak puanlayıcı sayısının bir arttırılması durumunda G katsayısındaki artma 0.056 olmaktadır. Benzer durum,  $\varphi$  katsayısı için de incelenirse, puanlayıcı sayısının sabit tutularak görev sayısının iki katına çıkarılmasında meydana gelen artma 0.092 iken, görev sayısı sabit tutularak puanlayıcı sayısının bir arttırılmasındaki artma ise 0.039 olarak gözlenmektedir. En yüksek G ve Phi katsayıları tablo 6’da görüldüğü gibi sırasıyla 0.722 ve 0.348 olup bu katsayılarla 5 puanlayıcının 10 göreve verdikleri puanlar doğrultusunda ulaşılabileceği gözlenmektedir. Literatürde istenilen G ve Phi katsayı değerlerine ulaşılabilmek için en az kaç puanlayıcı ve kaç görev gerekir? sorusuna yanıt aramak için ayrı bir hesaplama yapılmış , 9 puanlayıcı ve 15 görev için hesaplanan değer sırasıyla 0.894 ve 0.802 olarak bulunmuştur.



**Şekil 1: Voleybol Becerileri Performansının Değerlendirilmesinin K Çalışması ile Madde ve Puanlayıcı Sayıları Senaryolarına Göre G Katsayısı**



**Şekil2: Voleybol Becerileri Performansının Değerlendirilmesinin K Çalışması ile Madde ve Puanlayıcı Sayıları Senaryolarına Göre Phi Katsayısı**

Benzer durum,  $\varphi$  katsayısı için de incelenirse, puanlayıcı sayısının sabit tutularak görev sayısının iki katına çıkarılmasında meydana gelen artma 0.092 iken, görev sayısı sabit tutularak puanlayıcı sayısının bir arttırılmasındaki artma ise 0.039 olarak gözlenmektedir.

Tablo 5, Şekil 1 ve Şekil 2’de uygulamadaki görev ve puanlayıcı sayılarının azalması durumunda  $G$  ve  $\Phi$  katsayılarının giderek azaldığı, arttırılması durumunda ise  $G$  ve  $\Phi$  katsayılarının giderek arttığı açıkça görülmektedir. Tüm madde ve puanlayıcı sayılarına göre,  $G$  katsayısı  $\Phi$  katsayısından daha yüksek değerlere sahiptir. Tablodaki değerlere göre, hem  $G$  katsayısı hem de  $\Phi$  katsayısı için de görev sayılarını iki katına çıkarmanın, puanlayıcı sayısını bir arttırmaya göre daha büyük artışa sebep olduğu söylenebilir. Fakat bu artışın çokta yüksek olmadığı ve yeterli genellenebilirlik ve güvenilirlik katsayılarına ulaşamadığı dikkat çekmektedir.

## BÖLÜM IV

### SONUÇ VE ÖNERİLER

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

#### 4.1. SONUÇLAR

Araştırmanın bulgularına dayalı olarak ulaşılan sonuçlar aşağıda açıklanmıştır.

1) “Voleybol Becerilerine İlişkin Gözlem Formunun” klasik test kuramına göre iç tutarlılık düzeyinin belirlenmesi için Cronbach  $\alpha$  güvenilirlik katsayısı hesaplanmıştır. Tüm puanlayıcılar için ayrı ayrı hesaplanan Cronbach  $\alpha$  değerleri oldukça yüksek çıkmış ve puanlayıcıların her bir göreve verdikleri puan ortalamalarından elde edilen güvenilirlik katsayılarının, voleybol becerileri performansını değerlendirmede yer alan görevlerin birbirleriyle oldukça tutarlı olduğu başka bir ifadeyle iç tutarlığın sağlandığı sonucuna varılmıştır.

2) “Voleybol Becerilerine İlişkin Gözlem Formunun” değerlendirilmesinde yer alan dört puanlayıcı arasındaki tutarlılığın belirlenmesi için klasik test kuramına göre öncelikle parametrik olmayan istatistiksel bir teknik olan Kendall’ın Uyuşum Katsayısı kullanılmıştır. Ayrıca, her bir puanlayıcının verdiği puanlar ile diğer bir puanlayıcının verdiği puanlar arasındaki korelasyon katsayıları hesaplanmıştır. Yapılan bu analizlere göre, voleybol becerileri performansının değerlendirilmesinde yer alan dört puanlayıcının puanlamadaki düzeylerinin birbirinden farklı olduğu, puanlayıcıların verdikleri puan sıralamasının paralellik göstermediği, birbirleriyle tutarlı olacak şekilde puanlama yapmadıkları sonucuna varılmıştır.

3) “Voleybol Becerilerine İlişkin Gözlem Formu” nun değerlendirilmesinde yer alan 5 göreve tümüyle çaprazlanmış  $b \times g \times p$  modeli uygulanarak yapılan G çalışması sonucunda Voleybol becerileri performansını değerlendirme için kestirilen varyans ve toplam varyansı açıklama oranları incelendiğinde, birey (gp) etkisi için kestirilen

varyans bileşeninin toplam varyansın büyük bir oranını açıkladığı görülmüştür. Tek değişkenli modelle bireyler için kestirilen varyans bileşeni, toplam varyans içinde en düşük paya sahip varyans bileşenlerinden biri olmuştur. Bireylerin birbirinden ayrılması için kestirilen varyansın toplam varyans içindeki oranının en büyük olması istenilen bir durum olmasına karşın burada çok düşük çıkması bireyleri birbirlerinden ayıramadığının bir göstergesi olmuştur. Buna göre, yapılan voleybol becerileri performansının değerlendirmesinin, elde edilen boyutta bireyler arası farklılıkları ortaya çıkaramadığı sonucuna varılmıştır. Görev ana etkisi için kestirilen varyans, toplam varyansın küçük bir oranını (sıfıra çok yakın) açıklamaktadır. Buna göre, voleybol becerileri performansında yer alan görevlerin güçlük düzeylerinin birbirine yakın olduğu, bir görevin diğerinden daha zor olmadığı sonucuna varılmıştır. Puanlayıcı ana etkisi için kestirilen varyans bileşeninin toplam varyansın çok büyük bir oranı açıkladığı görülmüştür. Buradan, puanlayıcıların tüm sporcular için yaptıkları puanlamalar arasında farklılığın çok fazla olduğu, puanlamalar arasında bir tutarlılığın söz konusu olmadığı sonucu çıkarılmıştır.

4) Voleybol becerileri performansının değerlendirilmesinde tek boyut altında yer alan 5 görev sayısına ve bu maddeleri puanlayan 4 puanlayıcıya göre elde edilen genellenebilirlik (G katsayısı) 0.509 ve güvenilirlik ( $\varphi$  katsayısı) 0.208 olarak elde edilmiştir. Puanlayıcı varyansının toplam varyansa oranının yüksek olması sebebiyle, puanlayıcı sayısındaki artış ya da azalma genellenebilirlik ve güvenilirlik katsayılarını fazlaca etkilemektedir. Görev (madde) varyansının toplam varyansa oranı da puanlayıcı varyansına göre daha küçük olduğundan, görev sayısındaki artış ya da azalma genellenebilirlik ve güvenilirlik katsayılarını daha az etkilemektedir. Buna göre, genellenebilirlik ve güvenilirlik katsayılarının arttırılmasının istenildiği bir durumda beş puanlayıcı kullanılarak ve görev sayısını iki katına çıkararak (G katsayısı: 0.722 ve  $\varphi$  katsayısı: 0.348) var olan durumdaki (5 görev ve dört puanlayıcı) genellenebilirlik ve güvenilirlikten (sırasıyla; 0.509 ve 0.208) daha yüksek değerlere ulaşabileceği belirlenmiştir. Ancak bu durumda bile istenilen genellenebilirlik ve güvenilirlik katsayısına ulaşamadığı tespit edilmiştir. Voleybol becerileri performansının değerlendirilmesinde daha yüksek genellenebilirlik ve güvenilirliğe ulaşmak için

puanlayıcı ve görev sayılarının her ikisini birden arttırmanın gerekli olacağı sonucuna varılmıştır.

5. Birçok araştırmada geleneksel anlamdaki güvenilirlik ve geçerlik arasındaki farklılığı ortadan kaldırdığı için genellenebilirlik kuramı, tek bir çalışmayla güvenilirliğin ve geçerliğin belirlenmesini sağlayarak daha ekonomik ve pratik bir kuram olarak önerilebilir ifadesi yer almasına karşın yapılan bu araştırmanın sonucu bu öneriyi desteklememektedir. Klasik test kuramı ve genellenebilirlik kuramının benzer sonuçlar ürettiği görülmüştür.

## 4.2. ÖNERİLER

### *A.Voleybol Becerilerine İlişkin Gözlem formunun Değerlendirilmesine Dair Öneriler*

1.Voleybol Becerilerine İlişkin Gözlem formunun değerlendirilmesinde puanlayıcılar arası güvenilirlik çok düşük olduğundan puanlayıcı ve görev sayısını arttırarak istenilen güvenilirlikte ölçme yapmak daha uygun olabilir.

2.Voleybol Becerileri performansının değerlendirilmesinde puanlayıcıların, puanlama yöntemine ilişkin açık, anlaşılır ve yeterli derecede bilgilendirilmesi ya da uygulamalı olarak eğitilmesi, puanlayıcılar arası uyumun istenilen düzeye ulaşabilmesi açısından yararlı olabilir.

3. Puanlayıcıların uzmanlık alanları dikkate alınarak seçmek daha uygun olabilir.

4. Puanlayıcılar bilgi düzeyi daha homojen bir gruptan seçilebilir. Örneğin puanlayıcıların alt yapıda çalışmış beden eğitimi öğretmenlerinden seçilmesi uzmanlık alanlarını ön plana çıkarmayacakları düşünülerek puanlayıcılar arası tutarlılığı artırabileceği düşünülmektedir.

### *İleride Yapılabilecek Çalışmalar İçin Öneriler:*

1. Sporcuların maçlardaki performanslarını izleyip ona göre puanlama yapmak yerine, voleybol becerilerinin verili görevleri (örneğin her sporcunun 15 tane tekniğe uygun ve

hedefe yönelik pas atması, ölçütlerin belirlenmesi vb.) sırasında gözlenerek hem görevlerin hem de verili görevler sonrasında elde edilecek puanların birlikte değerlendirilebileceği bir ölçek geliştirilerek aynı tür bir çalışma daha yapılabilir. Ölçme aracındaki maddelerin belirlenmesi sürecinde puanlayıcılarında bulunması yararlı olabilir.

2. Voleybol Becerileri performansının puanlanmasında kullanılan gözlem formu yerine analitik rubrikle puanlama yapılarak benzer bir çalışma uygulanabilir.

3. Voleybol Becerileri performansının puanlanmasında hem holistik hem de analitik rubrik kullanılarak aynı çalışma üzerinde bu iki puanlama yöntemi karşılaştırılarak benzer bir çalışma uygulanabilir.

4. Voleybol Becerileri dışında farklı spor branşlarında uygulanan performans değerlendirmeleri için benzer güvenilirlik çalışmaları yapılabilir.

### ***B. Klasik Test Kuramı ve Genellenebilirlik Kuramı Kullanımına İlişkin Öneriler***

1. Performans değerlendirmeden elde edilen puanların hangi ölçek yapısına uygun olduğu, ölçmenin hangi amaçla yapıldığı ve elde edilen verilerin kullanılacak kuramın varsayımlarını sağlayıp sağlamadığı göz önüne alınarak analiz için hangi kuramın kullanılmasının uygun olacağına karar verilebilir. Başka bir ifadeyle hangi kuramda daha iyi işlediği belirlenebilir.

2. Bütün kuramların yararları göz önüne alındığında, yapılacak performans değerlendirmesi çalışmalarında sadece klasik test kuramına dayalı güvenirlüğün belirlenmesi yerine genellenebilirlik kuramının da beraberinde kullanılmasının yararlı olacağı söylenebilir.

3. Geçerlik ve güvenirlilik değerleri değiştikçe genellenebilirlik değişmektedir. Farklı güvenirlilik ve geçerlik katsayılarına sahip ölçekler kullanılarak elde edilen veriler ile aynı ölçeklerin genellenebilirlik kuramından elde edilen güvenirlilik katsayılarındaki tutarlılığa bakılarak, ölçmelerin psikometrik niteliklerindeki değişmeler araştırılabilir.

***İleride Yapılabilecek Çalışmalar İçin Öneriler:***

Bu çalışmada Voleybol Becerilerine İlişkin Gözlem formunun değerlendirilmesinin güvenilirliği genellenebilirlik kuramının b x g x p modeli ile analiz edilmiştir. Benzer çalışmalar genellenebilirlik kuramının farklı modelleri ve farklı değişkenlik kaynakları ele alınarak yapılabilir.

## KAYNAKÇA

- Açıkada, C. ve Ergen, E. (1990). *Bilim ve Spor*. Ankara: Büro-Tek Matbaacılık.
- Airasian, P.W., (2001). *Classroom Assesment*. Mc Graw Hill, Boston College
- Alharby, E. R. (2006). *A Comparison Between Two Scoring Methods, Holistic vs. Analytic Using Two Measurement Models, The Generalizability Theory and The Many Facet Rasch Measurement Within The Context of Performance Assessment*. Yayınlanmamış doktora tezi. The Pennsylvania State University.
- Ataay, İ.D. (1990). *İş Değerlendirme ve Başarı Değerlendirme Yöntemleri*. İstanbul İşletme Fakültesi Yayınları.
- 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.
- Balcı, A. (2001). *Sosyal Bilimlerde Araştırma: Yöntem, Yeti ve İlkeler*. Ankara: Pegem Yayıncılık.
- Baysal, M.E. (1999). *Veri Zarflama Analizi ile Orta Öğretimde Performans Ölçme*. Yayınlanmış Yüksek Lisans Tezi. Gazi Üniversitesi Fen Bilimleri Enstitüsü. Ankara.
- Baykul, Y. (2000). *Eğitimde ve Psikolojide Ölçme: Klasik Test Teorisi ve Uygulaması*. Ankara: ÖSYM Yayınları.
- Brennan, R.L. (2001) *Generalizability Theory*. Springer- Verlag. New York.
- Brookhart, S. M. (1999). *The Art and Science of Classroom Assessment: The Missing Part of Pedagogy*. ASHE-ERIC Higher Education Report (Vol.27, 1). Washington DC: The George Washington University. Graduate School of Education and Human Development.
- Budak, G.(1999). *İşletmeleri Başarıya Ulaştıran Yol Organizasyon Yapısı Birey İş Doyumu, Uyumu*. İzmir: Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi.

- Canman, D.(2000). *İnsan Kaynakları Yönetimi*. Ankara: Yargı Yayınevi.
- Crocker, L. M. ve Algina, L. (1986). *Introduction to Classical an Modern Test Theory*. New York: Holt, Rinehart and Winson.
- Cronbach, L. J. (1984). *Essential of Psychological Testing*. New York: Harper and Row Publish.
- Cronbach, L. J., Rajaratnam, N., & Gleser, G. C. (1963). *Theory of generalizability: A liberalization of reliability theory*. British Journal of Statistical Psychology. 16, 137-163.
- Çıtak Gözen, G. (2007). *Klasik Test ve Madde-Tepki Kuramlarına göre çoktan Seçmeli Testlerde Farklı Puanlama Yöntemlerinin Karşılaştırılması*. Ankara Üniversitesi Eğitim Bilimleri Enstitüsü. Ankara.
- Darst, P.W., Zakrajsek, D.B. ve Mancini, V. H. (1999). *Analyzing Physical Education and Sport Instruction*. Illinois: Human Kinetics Books
- Deliceoğlu, G. (2009). *Futbol Yetilerine İlişkin Dereceleme Ölçeğinin Genellenebilirlik ve Klasik Test Kuramına Dayalı Güvenirliklerinin Karşılaştırılması* Doktora tezi. Ankara Üniversitesi. Ankara
- Emberson, E. E. ve Steven, P. R. (2000). *Item Response Theory for Psychologists*. Lawrence Erlbaum Associates. Publishers. Mahwah. New Jersey. London
- Ergun, N.-Baltacı, G. (1997). *Spor Yaralanmalarında Fizyoterapi ve Rehabilitasyon Prensipleri*. Ofset Fotomat. Ankara, s.50.
- Erkal, E.M.( 1981 ). *Sosyolojik Açıdan Spor*. İstanbul.
- Erkuş, A. (2000). “Ölçüt-Dayanaklı Değerlendirme Bağlamında Geliştirilen Sınıflama Tutarlığı İndeksleri ve Bazı Sorunları”. Türk Psikoloji Yazıları. s: 1-21
- Fan, X. (1998). *Item Response Theory and Classical Test Theory: An Empirical Comparison of Their Item–Person Statistics*. Educational and Psychological Measurement. 58(3) 357-381.
- Fröhner, B. (1999). *Voleybol Oyun Kuramı ve Alıştırmaları*. Çev: A. Cengiz Bağırhan Yayınevi. Ankara

- Goodwin, L. D. (2001). *Interrater Agreement and Reliability*. Measurement in Physical education and Exercise Science. 5 (1), 13-14.
- Güler, N (2008). *Klasik Test Kuramı Genellenabilirlik Kuramı ve Rasch Modeli Üzerine Bir Araştırma* Doktora tezi. Hacettepe Üniversitesi: Ankara.
- Güler, N. (2009). *Genellenebilirlik Kuramı ve SPSS ile GENOVA Programlarıyla Hesaplanan G ve K Çalışmalarına İlişkin Sonuçların Karşılaştırılması* Eğitim ve Bilim, Cilt 34, Sayı 154
- Gürsakar N. (2001). *Bilgisayar Uygulamalı İstatistik-I*. İstanbul: Alfa Yayınları.
- Gürses, C. ve Olgun, P. (1989). *Sportif Yetenek Araştırma Metodu*. İstanbul: Türk Spor Vakfı. Araştırma No: 1.
- Hambleton, R. K., Swaminathan, H. ve Rogers, H.J. (1991). *Fundamentals of Item Response Theory*. Sage Publications. The International Professional Publishers.
- Hambelton R.K. ve Swaminathan H. (1985). *Item Response Theory: Principles and Application*, Kluwer: Nijhoff Publishing
- Howell, D. C. (2002). *Statistical Methods for Psychology*.(Fifth Edition). Thomson Learning Academic Research Center, USA.
- İskender, T. (2002). *Voleybolda 4 Numaradan Yapılan Smacın Elit ve Elit Olmayan Smaçörlerin Teknikleri arasındaki mekanik farklar* H. Ü. Voleybol Bilim ve Teknoloji Dergisi 9.29.sf. 3-6 Onay Ajans.
- Johnson, R.L., Penney, J. ve Gordon, B. (2000). *The Relation Between Score Resolution Methods and Interrater Reliability: An Empricial Study of an Analytic Scoring Rubric*. Applied Measurement in Education. 13(2),121-138
- Kantor, R., Lee, Y. W. (2007). *Evaluating Prototype Tasks and Alternative Rating Schemes for a New ESL Writing Test thtough G Theory*. International Journal of Testing. 7(4), 353-385.
- Kaynak, T. (1998). *İnsan Kaynakları Yönetimi*. İstanbul Üniversitesi İşletme Fakültesi Yayınları. Yayın No: 276, s.32- 35, İstanbul.
- Kıroğlu, G. (2001). *Uygulamalı Parametrik Olmayan İstatistiksel Yöntemler*. Ankara: Paymaş Yayınevi.

- 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.
- Klein, S. P., Stecher, B. M., Shavelson, R. J., McCaffrey, D., Ormseth, T., Bell, R. M. ve diğerleri. (1998). Analytic versus Holistic Scoring of Science Performance Tasks. *Applied Measurement in Education*. 11, 121-137.
- Kuter, M. , Öztürk, F. (1997). *Antrenör ve Sporcu El Kitabı*. Bağırhan Yayınevi. Ankara
- Lee, G. ve 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
- Linn, R.L., Gronlund, N.E., (1995). *Measurement and Assessment in Teaching*. New Jersey. Prentice-Hal
- Lord, F. M. ve Novick, M. R. (1968). *Statistical Theory of Mental Test Test Scores*. New Jersey: Addison-Wesley. Co.
- Ludlow, C. B. (2001). *Using Running Records As A Benchmark Reading Assessment: Reability In Assessing Reading Progress*. The Degree of Doctor. Brigham Young University, Provo.
- Magnusson, D. (1967). *Test Theory*. Stockholm: Addison-Wesley Publishing Company.
- Mehrens, W. A., Irvin J. L (1991). *Measurement and Evaluation in Education and Psychology*. Fourth Edition. New York: Holt, Rinehart ve Winston
- Mertler, C. A. (2001). *Designing Scoring Rubrics for Your Classroom*. *Practical Assessment Research and Evaluation*, 7 (25).
- Moon, S. Y. (1995). *Performance Assessment: Measurement Issues of Generalizability, Dependability of Scoring and Relative Information on Student Performance*. Yayınlanmamış doktora tezi. The Florida State University, Talhasse.
- Moskal, B. M. (2000). *Scoring Rubics: What, When and How?*. *Practical Assessment, Research and Evaluation*, 7 (3).
- Mushquash, C. and O'Connor, B. P. (2006). *SPSS and SAS Programs for Generalizability Theory Analysis*. *Behavior Research Methods*. 38 (3), 542-547.

- Nalbantoğlu, F. (2009). *Performans Ölçümlerinde Genellenebilirlik Kuramıyla Farklı Desenlerin Karşılaştırılması* Yüksek Lisans Tezi: Hacettepe Üniversitesi. Ankara.
- Nitko, A. J. (2001). *Educational Assessment of Students* (3. Baskı). Upper Saddle River, NJ: Merrill.
- Nunnally, J. ve Bernstein, I. (1994). *Psychometric Theory*. New York: McGraw Hill.
- O'Malley, J. M., ve Pierce, V. D. (1996). *Authentic Assessment*. Boston: Addison-Wesley Publishing Company Inc.
- Oosterhof, A.(1994).*Classroom Applications of educational Measurement*, Second edition. Macmillan Publishing Company. Newyork,
- Özcelik, D.A. (1998). *Test Hazırlama Kılavuzu*. ÖSYM Yayınları. Ankara.
- Rentz, J. O. (1987). *Generalizability Theory: A Comprehensive Method for Assessing and Improving the Dependability of Marketing Measures*. Journal of Marketing Research. 24(1), 19-28.
- Schermerhorn, J.R. (1989). *Management and Productivity*. Third Edition. Printed in the United States America.
- Sencer, M, ve Sencer Y. (1978). *Toplumsal Araştırmalarda Yöntembilim*. Ankara: Doğan Basımevi.
- Shavelson, R. J. ve Webb, N. M. (1991). *Generalizability Theory: A Primer*. Sage Publications, USA.
- Songur, H. M. (1995). *Mahalli İdarelerde Performans Ölçümü*. Ankara: Mahalli İdareler Genel Müdürlüğü Yayın No:6.
- Sub, A., Valiga, M. J. ve Goa, X. (1997). *Using Generalizability Theory to Assess the Reliability of Student Ratings of Academic Advising*. Journal of Experimental Education. 65, 4, 367-379.
- Szymanski, E. M., ve Linkowski, D. C. (1993). *Human Resource Development: An Examination of Perceived Training Needs of Certified Rehabilitation Counselors*. Rehabilitation Counseling Bulletin. 37 (2), 163-176.
- Şencan, H. (2005). *Sosyal ve Davranışsal Ölçmelerde Güvenirlik ve Geçerlik*. Ankara: Sözkese Matbaacılık.

- Taşdelen, G., Kelecioğlu, H. Ve Güler, N. (2010). *Nedelsky ve Angoff Standart Belirleme Yöntemleri ile Elde Edilen Kesme Puanlarının Genellenebilirlik Kuramı ile Karşılaştırılması* Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 1(1), 22-28.
- Tavşancıl, E. (2002). *Tutumların Ölçülmesi ve SPSS ile Veri analizi*. Ankara: Nobel Yayın Dağıtım Ltd. Şti.
- Taylor, G.R., (2003). *Informal Classroom Assessment Strategies For Teachers*. The Scarecrow Pres. Lanham. Maryland, and Oxford.
- Tekin, H. (2000). *Eğitimde Ölçme ve Değerlendirme*. 14. Baskı. Yargı Yayınları. Ankara.
- Tiryaki, E.S. (1999). *Yeni Başlayanlar İçin Voleybol Öğretimi*. Bağırğan Yayınevi, Ankara
- Tobar, D.A., Stegner, A. J. ve Kane, A.T. (1999) *The Use of Generalizability Theory in Examining the Dependability of scores on the Profile of Mood states*. Measurement in Physical Education and Exercise Science. 3(3), 141-156.
- Traub RE. (1994). *Reliability for the Social Sciences*. London: Sage Publications;
- Turgut, F. (1977). *Eğitimde Ölçme Ve Değerlendirme Metotları*. Nüve Matbaası. Ankara
- Turgut F. (1993). *Eğitimde Ölçme ve Değerlendirme Metotları*. 9 baskı. Ankara: Saydam Matbaacılık.
- Turgut, F. ve Baykul, Y. (1992). *Ölçekleme Yetileri*. Ankara: ÖSYM Yayınları.
- Vurat, M. (2000). *Voleybol Teknikleri*, Bağırğan Yayınevi. Ankara
- Welk, J. G. (2002). *Physical Activity Assessments for Health- Related Research*. United States: Human Kinetics Publishers Inc.
- Woods, R.H. (1997). *Human Resources Management*. Educational Institute, American Hotel and Motel Association. Second Edition.
- Yıldırım, C. (1999). *Eğitimde Ölçme ve Değerlendirme*. ÖSYM Yayınları. Ankara

- Yelboğa, A. (2007). *Klasik Test Kuramı ve Genellenebilirlik Kuramına Göre Güvenirliğin Bir İş Performansı Ölçeği Üzerinde İncelenmesi*. Yayınlanmamış doktora tezi. Ankara Üniversitesi. Ankara.
- Yurdugül, H. (2006). *Paralel, Eşdeğer ve Konjenerik Ölçmelerde Güvenirlik Katsayılarının Karşılaştırılması*. Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi. 39(1)15-37.

## EK 1

### VOLEYBOL OYNAYABİLME BECERİLERİNE İLİŞKİN GÖZLEM FORMU

**YÖNERGE:** Aşağıda Voleybolcuların voleybol becerilerine (Pas, Manşet, Smaç, Blok ve Voleybol (spor) zekası) ait beceri bildiren bir grup ifade verilmiştir. Formda bulunan ifadelerin her biri için her voleybolcuyu ayrı ayrı değerlendirerek puanlamanız beklenmektedir. İl karmaları müsabakalarını izlerken“Voleybol Becerilerine İlişkin Gözlem Formu” üzerindeki beceri ifadelerinden “Çok kötü”, “Kötü”, “Orta”, “İyi” ve “Çok İyi” seçeneklerinden uygun gördüğünüz birini X ile işaretleyerek değerlendiriniz. Sizden beklenen, her bir beceri ifadesi bakımından müsabakasını izleyeceğiniz voleybolcuyu değerlendirmenizdir.

- Eğer ilgili ifade de belirtilen beceriyi voleybolcuda hiç gözlemediyseniz, ya da çok kötü gözlemediyseniz **“Çok Kötü” ( Görülmedi)** kategorisini,
- Beceriyi oluşturan hareketlerin gösterilmesi ancak hareketin tamamlanmaması durumunda **“Kötü” ( Yetersiz)** kategorisini,
- Becerinin doğru yapılması durumunda **“Orta“ ( Yeterli )** kategorisini,
- Becerinin mükemmel yakın gösterilmesi/yapılması durumunda **“İyi” (Oldukça Yeterli )** kategorisini,
- Becerinin mükemmel gösterilmesi durumunda **“Çok İyi” (Mükemmel)** kategorisini işaretleyiniz.

Bütün uygulamaları dikkatlice izleyip işaretsiz hiçbir ifade bırakmayınız. Ayırdığınız zaman ve gösterdiğiniz işbirliği için teşekkür ederiz. İyi çalışmalar dileriz.

Ad Soyad:

Tarih:

Doğum Tarihi:

Seçici Kurul Üyesi:

Bölge/ Takım:

Bölge:

İmza:

	Görülmedi	Yetersiz	Yeterli	Oldukça Yeterli	Mükemmel
<b>PAS BECERİSİ</b>	<b>Çok Kötü</b>	<b>Kötü</b>	<b>Orta</b>	<b>İyi</b>	<b>Çok İyi</b>
	0	1	2	3	4
1.Topa vuruş noktası					
2.Topa temas anındaki bilek fleksiyonu					
3.Geriye pas					
4.Topun altına girebilmesi					
5.Vücut pozisyonu					
<b>MANŞET BECERİSİ</b>	<b>Çok Kötü</b>	<b>Kötü</b>	<b>Orta</b>	<b>İyi</b>	<b>Çok İyi</b>
	0	1	2	3	4
6.Topa vuruş öncesi pozisyonu					
7.Topun geliş yönüne göre vücut pozisyonu					
8.Vuruş anında bacaklarını kullanması					
9.Vuruş anında kollarını yönlendirmesi					
10.Vuruş anında vücut pozisyonu					
<b>SMAÇ BECERİSİ</b>	<b>Çok Kötü</b>	<b>Kötü</b>	<b>Orta</b>	<b>İyi</b>	<b>Çok İyi</b>
	0	1	2	3	4
11.Topu en yüksek noktada yakalaması					
12.Görerek istediği yere hucüm etmesi					
13.Topa vuruş yüksekliği					
14.Kolunu açması					
15.Bileğini kullanması					
<b>BLOK BECERİSİ</b>	<b>Çok Kötü</b>	<b>Kötü</b>	<b>Orta</b>	<b>İyi</b>	<b>Çok İyi</b>
	0	1	2	3	4
16.Blok yapacağı yerin tespiti					
17.Blok için kollarının hareketi					
18.Blok zamanlaması					
19.Blok takibi					
20.Blokta vücut pozisyonu					
<b>VOLEYBOL (SPOR) ZEKASI</b>	<b>Çok Kötü</b>	<b>Kötü</b>	<b>Orta</b>	<b>İyi</b>	<b>Çok İyi</b>
	0	1	2	3	4
21.Tüm temel teknikleri uygularken kararı					
22.Defansta duracağı yeri belirleyebilmesi					
23.Hücumda bloğu ve rakip defansı takibi					
24.Servisi amaçlı kullanabilmesi					
25.Oyun takibi					

## ÖZGEÇMİŞ

### Kişisel Bilgiler

Adı Soyadı : Meltem Emel ÖZTÜRK

Doğum Yeri ve Tarihi : ANKARA,1973

### Eğitim Durumu

Lisans Öğrenimi : ODTÜ Fen ve Edebiyat Fakültesi Matematik Bölümü

Yüksek Lisans Öğrenimi : H.Ü. Eğitimde Ölçme ve Değerlendirme Bilim Dalı

Bildiği Yabancı Diller : İngilizce

Bilimsel Faaliyetleri :

### İş Deneyimi

Stajlar :

Projeler :

Çalıştığı Kurumlar : 1995-1998: Ankara Karacan Dersanesi

1998-2000: Polatlı Anadolu Lisesi

2000-2002: Ankara Özel Yüce Fen Lisesi

2002-2003: Beypazarı Uruş İlköğretim Okulu

2003-2007: Ankara Yıldırım Beyazıt Anadolu Lisesi

2007- : Ankara Reha Alemdaroğlu Anadolu Lisesi

### İletişim

E-Posta Adresi : meltememel@gmail.com