

**T.C.
ANKARA ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
EĞİTİMDE PSİKOLOJİK HİZMETLER ANABİLİM DALI
ÖLÇME VE DEĞERLENDİRME BİLİM DALI**

**FUTBOL YETİLERİNE İLİŞKİN DERECELEME ÖLÇEĞİNİN GENELLENEBİLİRLİK
VE KLASİK TEST KURAMINA DAYALI GÜVENİRLİKLERİNİN
KARŞILAŞTIRILMASI**

DOKTORA TEZİ

GÖKHAN DELİCEOĞLU

**Doç. Dr. Nükhet ÇIKRIKÇI DEMİRTAŞLI
DANIŞMAN**

**Ankara
Mart-2009**

ANKARA ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜ'NE

Bu çalışma jürimiz tarafından Eğitimde Psikolojik Hizmetler Anabilim Dalı, Ölçme Değerlendirme Bilim Dalı'nda DOKTORA TEZİ ÇALIŞMASI RAPORU olarak kabul edilmiştir.

Başkan.....

Prof. Dr. Nizamettin KOÇ

Üye

Prof. Dr. Ezel TAVŞANCIL

Üye

Prof. Dr. Hülya AŞÇI

Üye

Doç Dr. Nükhet ÇIKRIKÇI DEMİRTAŞLI (Danışman)

Üye

Yrd. Doç. Dr. Ömay ÇOKLUK BÖKEOĞLU

Onay

Yukarıdaki imzaların, adı geçen öğretim üyelerine ait olduğunu onaylıyorum.

.../.../ 2009

ÖNSÖZ

Bu çalışmanın gerçekleşmesinde her aşamasında katkı ve desteklerini esirgemeyen tez danışmanım Doç. Dr. Nükhet ÇIKRIKÇI DEMİRTAŞLI'ya teşekkürlerimi sunarım.

Doktora öğrenimim ve çalışmam süresince farklı bakış açılarına sahip olmamı, yapıcı eleştirileriyle gelişmemi sağlayan Prof. Dr. Nizamettin KOÇ, Prof. Dr. Ezel TAVŞANCIL, Yrd. Doç. Dr. Ömay ÇOKLUK BÖKEOĞLU'na teşekkür ederim.

Doktora öğrenimim sırasında Ölçme ve Değerlendirme konusunda bilgilenmemi sağlayan Prof. Dr. Süleyman Çetin ÖZOĞLU, Prof. Dr. Yaşar BAYKUL, Doç. Dr. Şener BÜYÜKÖZTÜRK ve Yrd. Doç. Dr. Ömer KUTLU'ya teşekkürü borç bilirim.

Bu çalışmanın spor bilimleri bağlamında incelenmesi ve geliştirilmesinde yapıcı önerileri ve katkıları için Prof. Dr. Hülya AŞÇI ile araştırmanın yapılandırılmasında yorum ve desteklerini esirgemeyen Dr. Atilla YELBOĞA'ya teşekkürlerimi sunarım.

Çalışmam sırasında beni yalnız bırakmayan sevgili, arkadaşlarıma, öğrencilerime ve özellikle destekleri ile her zaman yanımda olan aileme minnettarım.

Gökhan DELİCEOĞLU

ÖZET

FUTBOL YETİLERİNE İLİŞKİN DERECELEME ÖLÇEĞİNİN GENELLENEBİLİRLİK VE KLASİK TEST KURAMINA DAYALI GÜVENİRLİKLERİNİN KARŞILAŞTIRILMASI

DELİCEOĞLU, Gökhan

Doktora, Ölçme ve Değerlendirme Bilim Dalı

Tez Danışmanı: Doç. Dr. Nükhet ÇIKRIKÇI DEMİRTAŞLI

Mart, 2009, 100 sayfa

Araştırmada, futbolcuların teknik yetilerinin tespit edilmesinde kullanılan Futbol Yetilerine İlişkin Dereceleme Ölçeği'nden elde edilen ölçümlerin; Klasik Test Kuramı (KTT) ve Genellenebilirlik (G) Kuramına dayalı olarak belirlenen güvenirlik katsayılarının düzeyleri saptanmış ve birbirleriyle karşılaştırılmıştır.

Çalışma grubunu, 2007-2008 yılları arasında Ankara ilindeki profesyonel spor kulüplerin futbol takımlarının alt yapılarındaki toplam 72 futbolcu (ortalama-yaş= 17,13±1,15 yıl; antrenman-yaş= 6,26±1,12 yıl) oluşturmuştur.

Futbolcuların teknik yetilerinin (pas, şut çekme, top sürme ve top kontrolü) ölçülmesi amacıyla 56 maddeden oluşan "Futbol Yetilerine İlişkin Dereceleme Ölçeği" (Luxbacher, 1991) kullanılmıştır. Dereceleme ölçeğine ait maddeler, beş dereceli ölçekte puanlayıcılardan (4 puanlayıcı), "Görülmedi", "Yetersiz", "Yeterli", "Oldukça Yeterli" ve "Mükemmel" seçeneklerinden birini işaretlemeleri istenmiştir.

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 güvenirliği için Cronbach α (alfa) katsayısı, puanlayıcılar arasındaki tutarlılık için Kendall uyum katsayıları hesaplanmıştır.

Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin alt boyutları ve ölçeğin toplamından elde edilen puanların güvenirliği için puanlayıcı sayılarının 3, 4 ve 5 olduğu ve madde sayılarının bir artırılıp, bir azaltıldığı koşullarda, G kuramında ana

ve ortak etkilerin varyans bileşenlerin kestirilmesi için MGENOVA paket programında 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_{xm}x_p ortak etkisi varyans bileşeni takip etmektedir. Bunların dışındaki temel ve ortak etkilere ait varyans bileşenlerinin ise sifıra yakın olduğu görülmektedir.

Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin puanlamasından elde edilen güvenilirlik katsayıları iç ölçütlere göre incelendiğinde G katsayısı ve Cronbach Alfa katsayıları beklenen değerlerinden yüksek olduğu görülmektedir. Phi katsayısı ile Kendall W Güvenirlik katsayılarının beklenen değerlerinden düşük olduğu görülmektedir.

Bu bağlamda, araştırmamızdaki potansiyel hata kaynaklarının fazla olduğu durumlara benzer çalışmalarda G Kuramı, KTK'na alternatif oluşturmaktadır. Bununla birlikte, Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin ve ölçeğe ait alt boyutların güvenilirliğinin yüksek olması, ölçeğin güvenilir bir ölçek olarak değerlendirebileceğini göstermektedir.

ABSTRACT

THE COMPARISON OF THE RELIABILITIES OF THE SOCCER ABILITIES' RATING SCALE BASED ON THE CLASSICAL TEST THEORY AND GENERALIZABILITY

DELİCEOĞLU, Gökhan

Ph. D. Department of Measurement and Evaluation

Advisor: Assoc. Prof. Nükhet ÇIKRIKÇI DEMİRTAŞLI

March, 2009, 100 pages

In the research, the measurements which were obtained from the Rating Scale related to the Soccer Abilities, used for determining the technical abilities of soccer players; Classical Test theory (CTT) and the levels of reliability coefficients stated at the base of the Generalizability (G) Theory were fixed and these levels were compared among them.

The Research group consists of the 72 soccer players (Mage= 17,13±1,15 years; Mtraining-age=6,26±1,12 years) who are playing in the Feeder Clubs of Professional Soccer teams in Ankara.

By the aim of the measuring technical abilities of the soccer players (pass, shooting, dribbling and control of ball), the rating scale which was composed of 56 items, was used (Luxbacher, 1991). It was demanded that the raters mark one of the options; “ not seen”, “deficient”, “sufficient”, “Considerably sufficient” and “Perfect” for the items belonged to the Rating Scale, in the five graded scale,

For the consistency between the pre-grading and final grading that were obtained from the Grading Scale related to soccer ability, Pearson correlation coefficient; for the internal consistency reliability of the elements, Cronbach α (alfa) coefficient; for the consistency among the raters, Kendall concordance coefficients were figured out.

For the reliability of the grades that were obtained from the entire and sub-dimensions of the Rating Scale Related to the Soccer Abilities, in the conditions of the slightly decrease and increase of the element numbers and existency of the

rates' numbers as 3, 4 and 5, in the G theory for estimating the variance components of main and common effects, the G and Phi coefficients are found in MGENOVA packet program.

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 he individual main influence. This is followed by bmxp common influence variance component. Except these, it is observed that the other variance components belonging to the main and common influences are closed to zero.

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

In this context, the G theory constitutes an alternative to the CTT in the similiar studies with the cases of the exceeding the potantial error sources in our study. Nonetheless, being high of reliability of the Soccer ability rating coefficient's joint scale and sub-dimensions of the scale indicated that the scale could be evaluated as a confidential scale.

İÇİNDEKİLER

Sayfa No

JURİ ÜYELERİNİN İMZA SAYFASI	i
ÖNSÖZ	ii
ÖZET	iii
ABSTRACT	v
İÇİNDEKİLER	vii
TABLolar LİSTESİ	x
BÖLÜM I	1
GİRİŞ	1
Problem	1
Yeti ve Futbol	2
Futbol'da Yeti Çeşitleri.....	3
Performans Kavramı	7
Performans Değerlendirme.....	8
Performans Değerlendirme Yöntemleri.....	8
Performansı Ölçme	9
Psiko-Motor Davranışların Gözlemlerle Ölçülmesi	11
Psiko - Motor Yetileri Puanlama Yöntemleri	12
Psiko - Motor Yetileri Ölçmede Kullanılan Araçlar	12
Ölçmede Hata Kavramı.....	15
Performansı Puanlamada Yapılan Hatalar	18
Puanlamada Yanlılık	19
Amaç	24
Önem	25
Sayıtlar	26
Sınırlılıklar	26
Tanımlar	26
Kısaltmalar.....	27

BÖLÜM II	28
KAVRAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR	28
Psikometride Kullanılan Başlıca Test Kuramları	28
I. Klasik Test Kuramı	29
Puanlama Güvenirliğini Kestirme Yöntemleri	35
II. Madde Tepki Kuramı	41
III. Genellenebilirlik Kuramı	43
Genellenebilirlik Çalışması	45
Karar Çalışmaları	50
İlgili Araştırmalar	54
Yurt İçinde Yapılan Araştırmalar	54
Yurt Dışında Yapılan Araştırmalar	56
BÖLÜM III	59
YÖNTEM	59
Araştırmanın Modeli	59
Çalışma Grubu	59
Veri Toplama Araçları	59
Verilerin Toplanması	61
Verilerin Analizi	62
BÖLÜM IV	65
BULGULAR VE YORUMLAR	65
İlk ve ikinci Puanlamaya ait Betimsel İstatistikler	65
İlk ve İkinci Puanlamalar arası Güvenirlik Katsayıları	66
Cronbach alfa İç tutarlılık Güvenirliği	69
Puanlayıcılar arası Güvenirlik	71
Dereceleme Ölçeğinden Kestirilen G ve Phi Katsayıları	72
Alt Boyutlara ait Kestirilen Varyans ve Toplam Varyansı Açıklama Oranları	74
Alt Boyutların Puanlamasına İlişkin Kestirilen G ve Phi Katsayıları	81
KTK ve G Kuramında Kestirilen Güvenirlik Katsayıları	85

BÖLÜM V	87
SONUÇ VE ÖNERİLER	87
Sonuçlar	87
Öneriler	89
BÖLÜM VI	91
KAYNAKLAR	91
EKLER. Futbol Yetisi Dereceleme Ölçeği	98

TABLolar LİSTESİ

Tablo No	Sayfa No
Tablo 1. G Kuramının bmxpx Deseninden Elde Edilen Puanlara Ait Eşitlikler	48
Tablo 2. İki Değişkenlik Kaynaklı Tesadüfi Desen için Varyans Bileşenlerinin Kestirilmesine Ait Eşitlikler	49
Tablo 3. Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin İlk ve İkinci Puanlamaya Ait Betimsel İstatistikler	65
Tablo 4. Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin İlk ve İkinci Puanlamalar Arası Güvenirlik Katsayıları.....	66
Tablo 5. Futbol Yetilerine İlişkin Dereceleme Ölçeğine İlişkin Dört Puanlayıcının İlk ve İkinci Puanlamalar Arası Güvenirlik Katsayıları	68
Tablo 6. Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin Cronbach Alfa İç Tutarlılık Güvenirliği	69
Tablo 7. Puanlayıcıların Puanlamalarına Ait Toplam Ölçek ve Alt Ölçekteki Cronbach Alfa İç Tutarlılık Güvenirliği.....	70
Tablo 8. Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin Dört Farklı Puanlayıcının Puanlamaları Arasındaki Kendall W Güvenirlik Katsayıları	72
Tablo 9. Futbol Yetilerine İlişkin Dereceleme Ölçeği Puanlamasının Çok Değişkenli K Çalışması Sonucunda Kestirilen G ve Phi Katsayıları	73
Tablo 10. Futbol Yetilerine İlişkin Dereceleme Ölçeğine İlişkin Puanlamada bmxpx Modeli İçin Alt Boyutlarına Ait Kestirilen Varyans ve Toplam Varyansı Açıklama Oranları	75
Tablo 11. Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin Alt Boyutları İçin K Çalışması Senaryolarına Göre Kestirilen G ve Phi Katsayıları.....	82
Tablo 12. Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin Puanlanmasından Elde Edilen Verilere Göre KTK Ve G Kuramından Kestirilen Güvenirlik Katsayıları..	85

BÖLÜM I

GİRİŞ

Bu bölümde araştırma problemi, amaç, önem, sınırlılıklar ve tanımlar ile ilgili açıklamalara yer verilmiştir.

Problem

Sporcuların, müsabaka sırasında yetilerini en üst düzeyde sergilemeleri gerekmektedir. Antrenmanlarla geliştirilmesi istenilen yetilerin futbol sürecinin başında ve sürece devam ederken geçerli ve güvenilir ölçme araç ve yaklaşımlarının kullanılması ile ölçülmesi, futbolcu hakkında verilecek kararlar bakımından önemlidir.

Dünyada olduğu gibi Türkiye’de de en çok ilgi gösterilen spor dalı, futboldur. Bu yönelimin getirisi çocuklar ile gençleri zararlı alışkanlıklardan uzak tutmak amacıyla spor kulüpleri bünyesinde alt yapı sporcuları yetiştirmektedir. Nitelikli futbolcuları yetiştirmek amacıyla kulüplerin alt yapılarında futbolcuların yetileri gözlemlenmektedir. Yetilerin değerlendirilmesinde yoğun olarak gözlem yaklaşımı kullanılmaktadır. Bu bağlamda en uygun gözlemin yapılabilmesi amacıyla, puanlayıcıların (antrenörlerin), futbolcuları puanlamaları ile elde edilen sonuçlarının tutarlılığının kuramsal olarak çalışılması, psikomotor yetilerin güvenilir biçimde ölçülmesi bakımından önemlidir.

Psikomotor açıdan hedeflenen bir yetinin 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. Futbolda yetilerin değerlendirilmesinde kullanılan ölçütlerin doğru ve yeterli olarak

belirlenmesi ve standart ölçme araçlarının oluşturulması, bunlara dayalı olarak isabetli değerlendirmelerin yapılabilmesi açısından önem taşımaktadır. Bu bağlamda, teknik ve futbol, performans, ölçme araçları, güvenirlik, puanlayıcı kavramları ele alınarak açıklanmaya çalışılmıştır.

Yeti ve Futbol

Yeti (aptitude) kavramı, bireyin hâlihazırda ve belli ölçülerde gelecekte yapabileceklerini belirler. Yetinin zihinsel dayanağa sahip olduğu ve böylece belli bir güdülenme ile bireyin, bir alanda öğrenme deneyimlerinden belli bir ölçüde ve düzeyde yararlanabileceği anlamını içerir (Luxbacher,1991).

Bireyin doğuştan getirdiği güç anlamında kullanılan yeti kavramına, özel yetenek, örneğin, müzik özel yeteneği, resim özel yeteneği olarak da yaklaşıldığını ve bu şekilde ölçüldüğü görülmektedir (Bompa,1983).

Ekblom (1994)'a göre yeti; belirli bir sportif hareketin amaca uygun ve ekonomik şekilde gerçekleştirilmesini anlatır. Yetide mükemmellik; en zor müsabaka koşullarında, sportif alıştırmaların hareket yapılarını ekonomik ve mükemmel bir şekilde yaparak en yüksek verime ulaşmaktır.

Futbolda yeti; başarıyı etkileyen en önemli faktörlerden biridir. Bir diğer deyişle; yeti, motorik özelliklerin bileşkesi olan kondisyon ve taktikle birlikte performansa etki eden faktörlerden biridir (Özkara, 2004).

Bu anlamda yeti, belli bir alandaki bilgi ve becerileri kazanmak için sahip olunan özelliklerdir. Sporun içindeki farklı branşlarda, özel yeteneklere sahip sporcuların ve bu branşlarda eğitim alacakların bazı yetilere sahip olması beklenir. Futbol, çeşitli fiziksel ve psikomotor özelliklerin ve yetilerin gerekli olduğu bir spor branşıdır. Bir futbolcunun, sahip olduğu yetileri “teknik” olarak geliştirmesi ve göstermesi beklenir (Ekblom,1994).

Futbolda Yeti Çeşitleri

Futbol yetisi, futbol oyununun gerektirdiği hareketleri o anın koşullarına uygun olarak hatasız ve zamanında yapabilmektir. Uygun yetinin seçimi ve yetinin uygulanmasına göre pas, top kontrol yetisi, top sürme ve şut yetileri şeklinde sınıflandırılan yetiler sırasıyla açıklanmıştır.

Pas yetisi

Pas yetisi, futbolcunun temel yetisi olup, takım içindeki otokontrolü sağlar ve takım oyunu anlayışının gelişmesine önayak olur. Pas yetisinin gerçekleştirilmesinde; pasın isabetinin, şiddetinin ve zamanın önemli olduğu görülmektedir. Pasın isabeti için ayağın içi ve dışının iyi kullanılması gerekliken, pasın başarılı olması için vücudun dengesi, destek ayağın vücuda olan yakınlığı ve topa vuruş noktası büyük önem taşır. Pasın şiddeti de rakip oyuncuyu saf dışı edebilecek hızda olmasına ve şiddetin ayarlanmasına bağlıdır. Pasın zamanı ise pasın hangi pozisyonda ve hangi zamanda alınacağına ya da verileceğine bağlıdır. Başarılı bir pasın gerçekleştirilmesi, pasın isabeti, şiddeti ve zamanının çok iyi ayarlanmasına bağlıdır (Luxbacher,1991).

Pas yetisi, uygulanış şekline bağlı olarak; ayak içi ve ayak dışı pas yetisi olmak üzere iki bölüme ayrılmaktadır.

a) Ayak içi pas

Futbolda en sık kullanılan ve isabet oranı en yüksek pas çeşidi "Ayak içi pas" tır. Uygulama başarısı en fazla olan pas çeşitlerinden birisidir. Pas sırasında destek ayağı topun yanına yerleştirilip, vuruş ayağın iç kısmıyla topun orta kısmına vurulması gerekmektedir. Vuruş hareketi ritmik bir salınımla kalça ve dizden yapılmalıdır. Vuruş sırasında ayak bileği mümkün olduğunca sabit ve sıkı tutulmalı ve vuruş ayağı, topun gidiş yönüne doğru devam etmelidir (Worthington,1974).

b) Ayak dışı pas

Ayak içi pas yetisinde, destek ayağı, topun yanına yerleştirilip vuruş ayağın dışıyla topun ortasına veya orta altına ayak bileği mümkün olduğunca sert ve sabit tutularak vurulur. Vuruş sırasında hareket kalça ve dizden ritmik bir salınımla yapılmalıdır. Ayak dışı pas uygulaması genellikle sporcuların isabetlilik ve kuvvetli vurmada sorun yaşadıkları bir pastır. Ayak dışı pasta isabetli, uzun ve sert atabilmek için çok fazla ayak dışı pas çalışmaları yapmak gerekmektedir (Luxbacher,1991).

Top sürme yetisi

Top sürme yetisi, saha boyunca topu kontrol altında tutabilmek için gerekli olan bir yetidir. Bu yeti, pas vermeden veya kaleye şut atmadan topla kısa ve uzun mesafe koşabilmeyi gerektirir (Reilly,1996; Worthington,1974).

Top sürme yetisi, futbolun şahsi ve yaratıcı bir parçası olduğundan ancak oyun koşulları içerisinde geliştirilir. Etkin bir şekilde top sürme, büyük oranda enerji kullanımı gerektirir. Diğer bir tanıma göre de saha boyunca koşarken, ayağın bir kısmı ile topu devamlı şekilde kontrol etme ve yuvarlama işlemidir (Herbin,1976; Reilly, 1996).

Top sürme yetisi, bir veya daha fazla rakibi saf dışı etmek için başvuru olan bir yetidir. Futbolcunun, ayağının değişik bölümlerini kullanarak yaptığı kısa vuruşlarla topu oyun alanından çıkarmadan değişik yönlere doğru götürmesi olarak da tanımlanmaktadır. Top sürmede amaç; futbolcunun, rakip oyuncuların baskısının söz konusu olduğu ve olmadığı durumlarda topu uzun süre ayağında bulundurması veya değişik yönlere doğru değişik tempolar ile götürerek uzun süre topla sahip olmasının sağlanmasıdır (Ferah,1992; Yılmaz,1992).

Bu tanımlar ışığında top sürme yetisi; kişinin en uygun zamanda, en uygun yetiyle futbol topunu belli bir noktadan belli bir noktaya taşıması ve böylece etkili bir top sürme yaparak en uygun zamanda ve en uygun yetiyle topun olumlu

kullanması sırasındaki süreçte topla yaptığı hareketler bütünü olarak tanımlanabilir.

Top kontrol yetisi

Top kontrol yetisi, topun ve futbolcunun pozisyonu ve kontrolden sonraki adımın planlanması açısından önemlidir. Top kontrol yetisinde, topu kontrol altına almada kullanılan vücut kısımlarına göre farklı sınıflandırmalar yapılabilmektedir. Top kontrol yetisi, uygulanış şekline bağlı olarak; ayak içi, ayak üstü, ayak dışı ve diz üstü kontrol yetisi olmak üzere dört bölüme ayrılmaktadır.

a) Ayak içi top kontrol yetisi

Yuvarlanmakta olan ya da yüksekten gelen topların kontrolü için uygulanan bu yetide, vücut pozisyonu iç vuruştaki pozisyonun aynısıdır. Topa temas sırasında bacak kalçadan itibaren geriye doğru hareket ettirilir ve top ayağın içinde tutulur; bu sırada hareketler oldukça yumuşaktır (Worthington,1974).

b) Ayak üstü top kontrol yetisi

Havadan ya da orta yükseklikten gelen toplara uygulanan bu yetide, topun sekmesine fırsat bırakılmadan kontrol altına alınır. Ayağın üstü ile kontrol yetisinde ayağın mümkün olduğunca açılması gerekir. Vücut destek ayağın üzerinde dururken diz hafifçe kıvrılır ve gevşek bırakılan ayak kalçadan itibaren mümkün olduğu kadar havaya kaldırılır. Top ayağa geldiği sırada karşılanır ve ayak yere indirilir. Ayağın sürati yere yaklaştıkça yavaşlarken, vücut biraz geriye gider ve kollar dengeyi sağlar (Luxbacher,1991).

c) Ayak dışı top kontrol yetisi

Topun baskın olan ayağın tarafından ve havadan geldiği durumlarda ayak içi kontrol yapılma olanağı yoktur. Bu gibi durumlarda uzakta olan ayak kullanılır. Baskın olan ayak, vücudu taşıyan ayağın önünde çaprazlama hareket ederek ayağın dışıyla topu kontrol eder. Bu kontrol yetisinde, vücudun ağırlığı destek ayak üzerindedir ve destek ayak dizden hafifçe bükülmelidir. Kontrol eden ayak

ise vücudun önünde çapraz bir biçimde sallanır. Top yere düşünce, bilekten aşağı doğru dönük olan ayağın dış kısmı ile temas edilir (Worthington,1974).

d) Diz üstü top kontrol yetisi

Top göğüs ve ayakla kontrol edilemeyecek kadar yüksekte gelirse top kontrolü diz üstü ile yapılır. Diz üstü kontrol sırasında oyuncu, topun geliş yönüne doğru dönerek beklerken vücut ağırlığı bir ayak üzerindedir. Diğer ayak ise kalçadan kaldırılarak diz üstü ve diz altı kısımları 50-60 derecelik açı oluşturacak biçimde kıvrılır. Topun geliş yönüne göre dizin pozisyonu ayarlanır ve yumuşak bir hareketle kontrol edilir (Luxbacher,1991).

Şut yetisi

Şut yetisi, ayağın herhangi bir kısmı ile topu göndermektir. Vuruşun topun hangi kısmına yapılacağına önemli olması kadar, isabetli vuruşun yapılması da etkili hücum açısından önem taşımaktadır (John ve Malcom, 1991). Şut yetisinde, ayağın hangi bölgesinin topa temas ettiğine göre farklı sınıflandırmalar yapılabilmektedir.

a) Ayak içi şut yetisi

Bu yeti uygulanırken, bacak dizden hafifçe kıvrılarak vücudun dengesi sağlanır. Topun hemen yanında duran ayağın ucu vuruş yönüne dönük iken, gövdenin üst kısmı topun üzerini örter. Topa vuruş yapacak olan ayağın ucu dışa doğru dönük ve yumuşak bir hareketle arkadan öne doğru sallanır. Vücudun ağırlığı destek ayak üzerine verilip topa vuruş yapılır ve vuran kişi vuruştan sonra topu izler (Reilly, 1996).

b) Ayak üstü şut yetisi

Topu uzağa göndermek için yapılan bir şut yetisi olup, vücut pozisyonu ayak içi üst vuruş pozisyonunun aynısıdır. Şutu yapacak olan ayak, topun üzerinde dik olarak tutulur. Ayağın ucu ise yere dönük olduğundan, ayağın üst yüzeyi topa tamamen temas eder (Luxbacher,1991).

c) Ayak dışı şut yetisi

Oyun içerisinde zor pozisyonlarda kullanılır. Vücut pozisyonu ayak içi şut yetisindeki, vücut pozisyonuyla aynı olup, kalça hafif öne doğru itilirken, ayağın ucu aşağı doğru dönük biçimde şutu çekecek olan bacak içe doğru çekilir. Topa dokunma sırasında ise ayak tamamen sabittir. Top oyuncunun önüne doğru süratle yuvarlandığında ve havadan gelen toplarda da dış vuruş kullanılır (John ve Malcom, 1991).

Futbol'da kullanılan bazı yetiler yukarıda tanımlanmıştır. Ancak yetiyi ölçmek için sporcunun bu yetiyi gerçekleştirme gerekmektedir. Bu bağlamda yetinin uygulanması yani performans ve performansın değerlendirilmesi kavramları aşağıda açıklanmaya çalışılmıştır.

Performans Kavramı

Performans, belli bir amaca yönelik olarak yapılan planlar doğrultusunda ulaşılan noktayı belirleyen bir kavramdır. Başka bir deyişle performans, bir işi yapan bir bireyin, bir grubun veya girişimin o işte amaçlanan hedeflere yönelik olarak varabildiği ve neyi sağlayabildiğinin nicel ve nitel olarak tanımıdır (Türker,1998).

Baysal (1999)'a göre performans kavramı, bir kişinin sahip olduğu potansiyel veya gerçek bilgi, yetenek ve kabiliyetlerini hedeflerine veya beklentilerine ulaşabilmek için ne ölçüde kullanabildiğini tanımlayan bir kavramdır. Bir başka ifadeyle, kişinin sahip olduğu kapasitenin, bir işi belli zaman dilimi içinde başarıyla tamamlamada kullanabilme yüzdesidir.

Bir diğer tanıma göre performans (edim) bireyin görevini gerçekleştirmek için yaptığı her davranıştır. Bu tanıma göre, performans, bireyin ne yapabileceğinden çok, ne yaptığına ilişkindir (Başaran, 1985).

Performans Değerlendirme

Bireyin nasıl bir edim elde ettiğini yargılama süreci olarak tanımlanan performans değerlendirme; genel olarak derecelendirme (rating), değerlendirme (evaluation) ve ölçme (appraisal) sözcükleri ile birlikte kullanılmaktadır. Alanyazında performans değerlendirme, performans yönetimi, başarı değerlendirme, tezkiye, sicil gibi tanımlamalara rastlamakla birlikte; uluslararası alanyazında yer alan "performance appraisal" veya "performance evaluation" terimlerinin karşılığı olarak kullanılmaktadır (Dağdeviren, 2005).

Ataay (1990), performans değerlendirmeyi genel anlamda, kişinin yeteneklerini, gizil güçlerini, iş alışkanlıklarını, davranışlarını ve benzeri niteliklerini diğer kişilerle karşılaştırarak değerlendiren sistematik bir süreç olarak tanımlamıştır.

Performans Değerlendirme Yöntemleri

Performans değerlendirme yöntemleri, görevin niteliğine ve değerlendirme sonuçlarının kullanılacağı alana göre farklılık göstermektedir. Değerlendirme yöntemleri alan yazında, a) Kişilerarası karşılaştırmalara b) Ortak performans kriter ve standartlarına ve c) Bireysel performans standartlarına dayalı yöntemler olarak da gruplandırılmaktadır (Beach,1980; Darst ve ark.,1989).

a) Kişilerarası karşılaştırmalara dayalı yaklaşım

Bireyin performansının bir veya daha fazla bireyle karşılaştırması ile elde edilen başarı düzeylerine göre sıralama yapılan bu yaklaşımda, bireyler, karşılıklı olarak birbirleri ile karşılaştırılırlar. Karşılaştırılan bireylerin sayısı fazla olduğu zaman, çok sayıda karşılaştırmanın yapılması ve her defasında bir kararın verilmesi gerekli olduğundan, bu yöntemin uygulanması oldukça uzun zaman almaktadır. Sistemin en büyük sakıncası da, sıralama içinde farklı noktalarda yer alan iki bireyin birbirlerinden ne kadar daha çok ya da daha az başarılı olduklarının bilinmemesidir (Beach, 1980; Turgut ve Baykul,1992).

b) Ortak performans kriter ve standartlarına dayalı yaklaşım

Bireyin performansını diğer bireylerden bağımsız olarak kendi görev tanım kapsamı içinde ele alır ve değerlendirir. Bu yaklaşımın kişilerarası karşılaştırmalara dayalı yaklaşımdan farkı, kullanılan değerlendirme ölçütlerinin birden fazla olması ve ortak olarak belirlenmiş olmasıdır (Helvacı, 2002).

Puanlayıcılar, ölçekte belirlenen boyutlar açısından bireyin nitelik ve davranışlarının hangi derece içine girdiğini saptayarak değerlendirmeyi yapar. Ayrıca her derecenin bir puan değeri bulunduğundan, işaretlenen dereceleme puanlarının toplanmasıyla başarı puanı bulunur (Aksoy; 1986).

c) Bireysel performans standartlarına dayalı yaklaşım

Bu yöntemde, bireyin gerçekleştirdiği hedeflere, gösterdiği sonuç ve çıktılara göre oluşturulan bireysel standartlarla değerlendirilmeleri amaçlanmaktadır (Darst ve ark., 1989).

Alanyazında tanımlanan performans değerlendirme yöntemlerinin uygun yapılabilmesi için performansı ölçmenin doğru yapılması gerekmektedir. Aşağıda performans ölçme kavramı açıklanmaya çalışılmıştır.

Performansı Ölçme

Performansın (Psiko- motor yetiler) önemli olduğu müzik, dans, cimnastik vb. gibi spor branşları geniş ölçüde vücut hareketlerine dayanır. Psiko-motor yeti, oldukça karmaşık davranışlar zincirinden oluşan ve psiko-motor yönü ağır basan yetiler olarak ifade edilmektedir. Piyano çalmak, yazı yazmak, bisiklet sürmek, yüzmek, voleybol oynamak, teraziyle kütle ölçmek, mikroskop kullanmak gibi yetiler, psiko-motor yetilerdir. Psiko-motor yetileri ölçmenin oldukça güç olduğu ve bu tür ölçmede kullanılacak araç ve yöntemlerin geliştirilmesinde yetersizliklerin bulunduğu görülmektedir (Turgut, 1993).

Psiko-motor yetiler olarak tanımlanan spor branşlarında 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 özellikler saptanarak, fiziksel yetenek ve spor branşı arasındaki ilişkilere göre belirlenen ve yaşa bağlı olarak değişmeyen ideal sporcu modellerinin araştırılması özellikle son yıllarda gündemini korumaktadır (Gürses ve Olgun, 1989).

Performans ölçmenin sporcu başarısı için öneminin artması, bu kavram ile ilgili çalışmaların artmasına neden olmaktadır. Bu araştırmalardan elde edilen genel yöntemlere bakıldığında;

Performansın (Psiko-motor yetiler) ölçülmesinde şu üç genel yöntemden birine başvurulabilir;

1. Birey, hedef alınan yetiyi uygular; puanlayıcı, uygulama sırasında bireyin hareketlerini gözler ve puan verir. Bu yöntem, aslında, davranışın gözlenip ölçülmesidir.

2. Birey, hedef alınan yetiyi uygulayarak bir yeti ürünü ortaya çıkarır; Puanlayıcı, bu ürünü inceler ve ona puan verir. Bu yöntemde ürüne bakılarak bireyin yetisi hakkında bir değer yargısına varılır.

3. Birey, hedef alınan yetiyi uygulayarak bir ürün meydana getirir. Puanlayıcı, hem bireyin davranışlarını hem de ürünün oluşmasını, çeşitli basamaklarda gözler ve her basamakta ara puanlar verir. Bu işlemin sonunda da bir toplam puan elde edilir. Bu yöntemlerden hangisinin kullanılacağı, hedefin mahiyetine ve ölçmenin amacına bağlıdır (Turgut, 1993).

Psiko-Motor Davranışları Gözlemlerle Ölçme

Psiko-motor yetiler, karmaşık davranışlar zincirinden oluşur. Bir yetinin ölçülmesi için, bu zincirin önemli halkaları gözlenebilmeli ve gözlem sonuçları sayısallaştırılmalıdır. Psiko-motor yetiyi oluşturan davranışları gözleyerek ölçmede izlenecek yöntemin ana basamakları aşağıdaki şekilde olmalıdır:

- Yetiyi oluşturan davranışlar analiz edilmeli, bunlar arasından ölçme için gerekli kritik davranışlar saptanmalıdır.
- Kritik davranışların gözlenmesini ve puanlanmasını kolaylaştıracak bir araç hazırlanmalıdır.
- Bireyin, saptanan kritik davranışları uygulayacağı gözlem ve sınav durumları oluşturulmalıdır.
- Birey, oluşturulan bu durumlarda davranışı uygularken gözlenmeli gözlenen davranışlar puanlanmalıdır (Turgut, 1993).

Psiko-motor yetileri 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. Yetiyi oluşturan davranışlar zincirinde, ancak kritik davranış halkaları gözlenip ölçülmelidir (Darst ve ark., 1989).

Kritik davranışların gözlenip puanlanması için bir araç hazırlama işi, hedefe ait davranışların analizi iyi yapılırsa kolaylaşır. Analiz sonucunda seçilen kritik davranışlar, kullanışlı bir düzenle araçta yer almalıdır. Bir yetinin uygulanmasındaki basamaklara göre sıralama, çoğu halde kullanışlı 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 yetileri gerçek bir maçta gözlenebileceği gibi test için düzenlenmiş bir oyunda da gözlenebilir (Turgut, 1993,163).

Oluşturulan test durumunda bireyin kritik davranışlarının gözlenmesi, oldukça zor bir iştir. Ölçmede kullanılacak araç hazırlandıktan sonra, puanlayıcıların bu aracı kullanması için yetiştirilmesi yararlı olabilir. Psiko-motor yetileri ölçme işlemlerinde, yetiler 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 yeti tekrarlatılır ve her seferinde belli bir kritik davranış gözlenir (Welk, 2002).

Psiko - Motor Yetileri Puanlama Yöntemleri

Bir yeti çok küçük davranış birimlerine ayrıştırılmış ise, 0 ve 1 gibi bir puanlama yöntemi yeterli olabilir. Fakat yetiyi oluşturan davranışlar daha karmaşık ise ve derecelenebiliyorsa, daha fazla bölmeli bir ölçek kullanmak gerekir. Aksi halde ölçmenin duyarlılığı azalır ve puanların güvenilirliği düşer. Bazı yetiler, nesnelere bilinen büyüklükleriyle ölçülebilir. Örneğin, koşularda süre, dakika - saniye - salise olarak, atlamalarda, uzunluk ve yükseklik metre-santimetre olarak ölçülür (Darst ve ark., 1989).

Bazı durumlarda ise davranışın tekrar sayısı puan olarak alınabilir. Voleybol müsabakasında bir oyuncunun kaç defa servis attığı ve bunlardan kaç tanesini kaçırdığı (dışa attıkları ve ağa taktıkları) sayılabilir. Buna benzer yöntemlerle yetinin ölçülmesinde kullanılacak puanlama yönteminin, hangi amaçla kullanılacağı da dikkatle ele alınmalıdır. Puanlama yöntemi, değerlendirmenin amacına hizmet edecek ayrıntıda puanlar vermelidir (Welk, 2002).

Psiko - Motor Yetileri Ölçmede Kullanılan Araçlar

Psiko-motor yetileri ölçmek için, işaretleme listesi ve dereceleme tipi ölçme araçları geliştirilmiştir.

a) İşaretleme Listesi Tipi Araçlar

Bu tip araçlar, gözlenecek davranışın varlığını “x” veya “evet” gibi bir işaret veya kelime ile gösterilmeye imkân 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).

b) Dereceleme Ölçeği

Yapılandırılmış gözleme dayalı ölçmelerde testlerin puanlanması olarak tanımlanan, dereceleme ölçeği gözlenen yeti veya davranışın mükemmelliğini de göstermeye imkân vermektedir. Bu tür ölçeklerde puanlayıcının yargısı önemlidir. Puanlama, bir süreklilik üzerindeki durum konusunda bireyin öznel yargısı ve ölçek yolu ile bu yargının açık bir derecelendirmeye dönüştürülmesi şeklinde iki aşamalı bir süreçtir: Bunlardan birinci aşama için şu varsayım esas alınır. Puanlayıcılar istenen boyutun farkındadırlar ve onların bu boyut-süreklilik üzerinde kendi durumlarına karşı gelen uyarının öngördüğü derecelendirmeyi yapabileceği varsayılır. Puanlama sürecinin ikinci aşamasında puanlayıcının kendi referans çerçevesini kullanarak derecelendirme yapması beklenir. Referans çerçevesi de o anda derecelendirilen objeler setine, puanlayıcının geçmişte hatırladığı objeler setine ya da onun kendi tutumu gibi özel referans noktalarına bağlı olabilir (Balci, 2001).

Bireylerin ya da sporcuların eğitim ortamında edindikleri performansa dayalı bilgileri ölçmek için kullanılan geleneksel yöntemlerden farklı olarak “Tamamlayıcı ölçme-değerlendirme yöntemleri” geliştirilmiştir (Mehrens, 1992).

Tamamlayıcı Ölçme ve Değerlendirme Yaklaşımları

Tamamlayıcı ölçme-değerlendirme yöntemleri ile yapılan ölçmelerde puanlamanın güvenilirliğinin arttığı, performans değerlendirilirken ölçütler önceden belirlendiği için puanlayıcılar arasındaki tutarlılığının da arttığı ve daha güvenilir

sonuçların elde edildiği belirtilmiştir (O'Malley ve Pierce, 1996).

Tamamlayıcı ölçme-değerlendirme yöntemlerinden bazıları aşağıda açıklanmaktadır.

Dereceli Puanlama Anahtarı (Rubric)

Dereceli puanlama anahtarı, gözlemlere ait puanları tanımlanmış kategorilerden uygun düşen boyuta kaydetmeyi sağlayan bir ölçme aracıdır. Ölçütlere dayalı dereceli puanlama anahtarı ya da puanlama cetveli olarak da isimlendirilen dereceli puanlama anahtarları, gözlemlerin puanlanmış kategoriler halinde kaydedilmesini sağlayan bir çeşit araçtır. Puanlama anahtarları, kolayca gözleyebilecek somut davranışlardan çok, yazının kalitesi, cümlelerin akıcılığı, konunun ele alınışındaki bütünlük gibi daha çok soyut olan ve ancak uzman kişiler tarafından gözlenebilecek davranışların puanlanmasında kullanılırlar (Darst ve ark., 1989; Haladyna, 1999).

Eğitim sürecindeki ölçmelerde kullanılan derecelendirme ölçek çeşitleri aşağıda açıklanmaya çalışılmıştır.

a) Grafik derecelendirme ölçekleri

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 bunların işlevi, puanlayıcıya süresiz kategoriler sağlamak değil onun tercihini yerleştirmesine rehberlik etmektir. Etkili kullanım için uç ifadelerin aşırı olmasından kaçınılmalıdır (Balci, 2001).

b) Maddeli derecelendirme ölçekleri

Spesifik kategori ölçekleri olarak da isimlendirilen bu tür ölçeklerde puanlayıcı ölçek pozisyonlarınca düzenlenmiş kategorilerden birini seçer. Ölçek pozisyon ya da kategorilerinin sayısı, araştırma problemi ve yargıya göre değişmekle beraber tipik olarak 5-11 arası kategori bulunur. Kategorileri gösteren

sözel betimlemeler kategoriye uygun davranışı betimleyen kısa ifadeler şeklinde olmalıdır. Kategorilerin tanımlanmış olması puanlayıcılar arası değişkenliği azaltarak güvenilirliği arttırmaktadır (Johnson ve ark., 2000).

c) **Karşılaştırmalı derecelendirme ölçekleri**

Bu tür ölçeklerde puanlayıcıdan karşılaştırma yapması beklenir. Ölçekteki derecelendirme pozisyonları belli bir sosyal grup ya da evren temeline göre; yani belli özellikleri bilinen insanlara göre tanımlanır. Öğrencilerin sınıfta birbirlerine göre başarılarının değerlendirilmesi, bir öğrencinin sınıftaki arkadaşlarının yüzde kaçından daha başarılı olduğunun tespit edilmesi için bu tür ölçekler kullanılabilir. Bir kişiden sınıftaki bir öğrencinin liderlik yetisinin, A, B ya da C kişisinden hangisine daha yakın olduğunu derecelendirmesi istenebilir (Balci, 2001).

d) **Kategorik puanlama ölçekleri**

Likert ölçekleri olarak da isimlendirilen, kategorik puanlama ölçekleri derecelendirilmiş puan seviyelerinden oluşmaktadır ve en yaygın kullanılan puanlama ölçekleridir. Bu ölçekte, her puan kategorisinin ne anlama geldiği daha belirgindir ve ara puanlamalardan kaçınılmıştır. Kategorik puanlama ölçeklerinde kullanılacak seviye sayısı 3 ile 7 arasında değişir (Haladyna, 1999).

Ölçme aracı veya ölçme aracı dışındaki unsurlar ölçme sonuçlarını etkilemektedir. Bu bölümde ölçme sonuçlarını etkileyen ölçmede hata kavramı açıklanmaya çalışılmıştır.

Ölçmede Hata Kavramı

Eğitim ve psikoloji alanında yapılan ölçmeler öğrencilerin doğrudan gözlenemeyen özelliklerini kapsadığından dolayı psikofizik uzayda tanımlıdır. Psikofizik uzayda yapılan ölçmeler ise iki boyutta ele alınır. Bunlardan ilki; ölçülmeye çalışılan özelliğin “gerçek boyutu”, diğeri ise bireyin gözlenebilen davranışlarının puanlayıcı tarafından “algılanan” boyutudur. Örneğin i ve j gibi iki ölçme maddesi ele alındığında bireyin ölçülmek istenilen özelliğe ilişkin gerçek

puanı ile gözlenen puanı arasındaki farklılık hata olarak adlandırılır (Osburn, 2000).

Ölçmelere karışan çeşitli hatalar nedeniyle ölçülen özelliğin kestirilen gerçek değeri ortaya konulamaz. Üzerinde ölçme yapılan değişkenin gerçek puanı, ölçme aracı, ölçmeci ve ölçme ortamı gibi değişkenlerden kaynaklanan hatalardan etkilenmektedir. Ölçmelerde hata, ölçülmek istenmeyen miktarın ölçme sonuçlarına karışmasıdır. Ölçme ve ölçme sonuçlarına, sabit, sistematik ve tesadüfi hata olarak üç türde hata karıştığı görülmektedir. Sabit hatalar, her bir ölçmede aynı yönde etki yapan ve miktarı değişmeyen hatalardır. Sistematik hatalar belirli bir yönde ilerleyici veya gerileyici miktarlarda ortaya çıkarlar. Ölçmelerde esas inceleme konusu olan tesadüfi hatalar ise şansa ortaya çıkan ve hangi yönde etki ettiği kestirilemeyen hatalardır. KTK bu tür hataların belirlenmesi (kestirimi) üzerine kurulmuştur (Crocker ve Algina, 1986; Turgut, 1993).

Bir ölçme durumunda, elde edilen gözlenen puanlara hata karışması nedeniyle ölçülen özelliğe ait gerçek değer doğrudan elde edilmesi olanaklı değildir. Ölçme işlemi ile elde edilen gözlenen puan, gerçek puan ve hata puanından oluşur. Ölçme yoluyla elde edilen gözlenen puanlarla, ölçülen özelliğin gerçek değerine ulaşılması beklenilir. Ölçme yoluyla gerçek puana ulaşılması demek, ölçme ile elde edilen puanların hatasız olması anlamına gelir. Ölçme işleminde kullanılan ölçme aracı ne kadar hassas olursa olsun, bütün ölçme sonuçlarına farklı kaynaklardan karışan hataların olması kaçınılmazdır. Bu nedenle, ölçülen özelliğin gerçek değerine ölçme yoluyla doğrudan ulaşılması söz konusu değildir (Baykul, 2000).

Ölçme sonucunu tamamiyle hatalı kabul etmek mümkün olmadığı gibi, tamamiyle hatasız kabul etmek de mümkün değildir. Doğa bilimlerinde bile somut ve kararlı özelliklerin ölçülmesinde bir miktar hata vardır. O halde herhangi bir ölçme sürecinden elde edilen ölçme değerinde bir miktar gerçek ve bir miktar da

ölçme hatası vardır. Bu durum ölçmenin gerçek ve hata olmak üzere iki öğeden oluştuğunu göstermektedir (Carmines ve Zeller,1982; Dawson ve Trap, 2001).

Ölçme aracından standart koşullarda elde edilen ölçmelerdeki toplam değişkenliği oluşturan kaynaklardan biri ölçme hatasıdır. Ölçme hatasının toplam değişkenlik içindeki payının artması, toplam değişkenliğin değişmemesi koşuluyla öteki payları, sınırlayacaktır. Hatanın katkısının artması, ölçülmek istenen özellik bakımından kişiler arası gerçek farklardan gelecek değişkenliğin azalmasına yol açacaktır (Tavşancıl, 2002).

Ölçmelerin kararlılığının yorumlanmasında, iki farklı yaklaşım vardır. Birinci yaklaşım, birbirini izleyen ölçmelerde, bireyin grup içindeki sırasının değişmezliği; ikinci yaklaşım ise birbirini izleyen ölçme sonuçlarındaki hataların büyüklüğü, yani ölçüğün tekrarlı ölçmelerde aynı denekten yaklaşık olarak aynı ölçme sonucunun elde edilmesi ve dolayısıyla ölçmenin standart hatasının düşük olması ile ilgilidir (Thorndike ve ark.,1991).

Ölçmenin standart hatası, bireysel ölçmelerde görülen ölçme hatalarının büyüklüğü ile ilgilidir. Objenin çok sayıda (sonsuz) ölçüldüğü ve bu ölçmelerin sonuçları incelendiğinde, her bir ölçme, birbirinden az ya da çok farklı olacak ve bu ölçmeler bir dağılım gösterecektir. Bu dağılımın ortalaması bireyin ilgilenilen özellikle ilgili gerçek puanın tahminini verir. Standart sapması da ölçmenin standart hatası adını alır. Elde edilen puan, güvenilirlik katsayısı ve standart sapması biliniyorsa hataların değişkenliği başka bir deyişle, ölçmenin standart hatası kestirilebilir (Tavşancıl, 2002).

Crocker ve Algina (1986) ise ölçmenin standart hatasını; bir bireyin sonsuz sayıda tekrarlı ölçmelerinde gözlenen ve gerçek puanı arasındaki uyumsuzluğun standart sapması olarak tanımlamıştır.

Farklı psikometrik değerlendirmelerin kullanılmasında iki temel problem

bulunmaktadır. Bunlardan birincisi, deęerlendirmede kullanılan ölçeklerin geçerlik ve güvenilirliklerinin gösterilmesinin güçlüğüdür. Güvenilirlik için puanlama tekrar puanlama yaklaşımının kullanılmasında, ölçmeler arasında performansta oluşan gerçek bir deęişimin ölçme hatası olarak yorumlanma olasılığı bulunmaktadır. İkinci problem ise, psikometrik deęerlendirmelerin, deęerlendirme doğruluğunun doğrudan bir göstergesi olmamalarıdır. Deęerlendirmede kullanılan ölçeklerin güvenilir ve geçerli olduklarının gösterilmesi, bu ölçekler kullanılarak yapılan deęerlendirmelerin "doęru" ya da "yansız" olacağını garanti etmemektedir (Murphy ve Cleveland, 1995).

Puanlamada yansızlık önemlidir. Bununla birlikte bu ve benzeri durumlar puanlamalara hata olarak karışabilir. Puanlama sırasında yapılan hatalar aşağıda verilmiştir.

Performansı Puanlamada Yapılan Hatalar

Performansı puanlama öznel ve çok çeşitli faktörler tarafından etkilenmektedir. Puanlayıcıların, performans deęerlendirmelerini adil bir biçimde yürütme ve eksiksiz bir biçimde gerçekleştirmede hata ihtimalini artıran faktörleri tanıması, bu tür hataları en aza indirir. Performans deęerlendirmede, bireye yönelik puanlamaların beraberinde yanılıęı paylarını da taşıdıkları bilinmektedir. (Welk, 2002).

Psikolojik testlerin uygulandıęı koşulların ve bireyin teste karşı olan tutumu test puanlarının güvenilirliğini etkileyeceęi düşünölmektedir. Psikolojik testin, yeterlilięi olmayan puanlayıcılar tarafından uygulanması, testin uygulama yönergesinin izlenmemesi, test uygulama koşullarının uygun olmaması ve test uygulanan bireyin, testi cevaplama davranışında, "kasıtlı yanıtma" göstermesi test sonuçlarının güvenilirliğini düşürecektir (Cronbach, 1984).

Bu bağlamda, puanlamada yanılılık, test sonuçlarının güvenilirliğini düşüren

en önemli faktör olduğu görülmektedir.

Puanlamada Yanlılık

Yanlılık, ölçme sonucunda belli bir yönde ortaya çıkma eğilimdeki sistematik hatadır. Yanlılık, testi alan kişilerin belli bir alt kümesinin gereğinden az ya da fazla puan almasıdır (Haladyna, 1999).

Puanlayıcılar arasındaki tutarlılık, alanyazında değerlendiriciler arası tutarlılık, hakemler arası tutarlılık, kayıtçılar arası tutarlılık şeklinde farklı isimlerle anılmaktadır. Puanlayıcılar, belirli bir olguyla ilgili olarak puan veren veya değerlendirme yapan kişilerdir. Puanlayıcılar, yapılan ölçmelerde benzer puanları vermişlerse sonuçlar güvenilir demektir (Szymanski ve Linkowski, 1993).

Puanlayıcılar, değişik nedenlerle puanlama yapabilirler. Örneğin doktorlar hastalarını, yöneticiler personelini, antrenörler futbolcularını ve jüri üyeleri cimnastikçileri puanlamaktadırlar. Puanlayıcılar arasındaki puanlamaların güvenilirliği, ölçmeden elde edilen puanlar arasındaki tutarlılığa bağlıdır. Puanlayıcılar arası güvenilirlik "farklı puanlayıcılar tarafından yapılan puanlamaların puan ortalamalarındaki sapma derecesidir (Tinsley ve Weiss, 1975).

Puanlayıcıların aşırı uçları puanlamadan kaçınmaları bu uçların daha az aşırı yapılmasıyla ortadan kaldırılabilir. Bazen tepki kategorilerinin isimlendirilmesi (labeling) cömertlik etkisini artırabilmektedir. Bir bireyi derecelendirirken beşli bir ölçekte "orta" orta kategori olarak kullanılırsa, genelde "iyi" ve "pekiyi" üst iki kategori olarak kullanılma eğilimindedir. Bu yüzden orta kategori olarak "iyi" nin kullanılması süper, pekiyi, orta ve kötü gibi olması durumunda daha uygun olabilmektedir (Balcı, 2001).

Dereceleme ölçekleri bireylerin öznel puanlamalarına dayalı olduklarından

tesadüfî ve sistemli hata kaçınılmaz olur. “Halo etkisi”, “Abartma-cömertlik hatası” ve “Zıtlık hatası” gibi bazı sistemli hata türlerinde puanlayıcının yanlış puanlama yapması ölçme sonuçlarını olumsuz etkileyecektir. Puanlamada yorgunluk, dikkatsizlik, ilgi azlığı gibi tesadüfî hatalar da yapılabilir. Bu tür hatalar basit hatalar olup yetiştirme ve güdülenme bunların giderilmesinde rol oynayabilir. Ayrıca birden çok puanlayıcının kullanılması ve her bir puanlayıcının, puanlamalarının ortalamasının alınması tesadüfî hata etkisini azaltabilir. Böylece bağımsız hatalar ortalama alınarak birbirlerini dengeleyebilirler (Johnson ve ark., 2000).

Ölçme araçlarının güvenilirliğinin sağlanması, ölçme hatalarının kaynağının bilinmesine bağlıdır. Güvenirlikle ilgili alanyazın incelendiğinde, güvenilirliği etkileyen hataların nereden kaynaklandığına bağlı olarak, aşağıdaki gibi sistematik bir sınıflama elde edilebilir (Erkuş, 2003).

a) Özelliği ölçülen kişiler

Özelliği ölçülen kişilerin oluşturduğu örneklemin, ölçülen nitelik bakımından homojen olup olmaması güvenilirliği etkiler. Bununla birlikte; bireylerin, ölçmenin yapıldığı süre içinde buldukları özel koşullardan ya da geçirdikleri olumsuz yaşantılardan dolayı gerçek performanslarını ya da durumlarını ortaya koyamamaları, olması gerekenden ölçme sonuçlarının farklılaşmasına neden olacaktır (Crocker ve Algina, 1986).

b) Ölçmenin yapıldığı koşullar

Güvenilir bir ölçme aracı olsa bile standart koşullar altında yapılmayan bir ölçme, bireyler arasında ortaya çıkan farklılıkların, bireylerin gerçek puanlarındaki farklılıklardan kaynaklandığı şeklinde yorumlanabilir (Baykul, 2000).

c) Ölçme aracı

Ölçme araçları ölçülen davranış alanını çok iyi örneklemelidir. Bunun için davranış alanının tanımı çok iyi yapılmalı, ölçülecek davranış hedefleri çok iyi belirlenmeli, madde tipi, madde sayısı ve uygulama zamanı uygun biçimde seçilmelidir. Bir ölçme aracının güvenilirliği, elde edilen verilerin “uygunluk”, “anlamlılık” ve “faydalılık” olarak ifade edilebilmesidir (Baykul, 2000; Özen ve ark., 2006).

d) Ölçmeyi yapan kişi ya da kişiler

Performansa dayalı ölçmelerde yetenek kestirimini etkileyen faktörlerden biri de puanlayıcıların katı ya da ılımlı puanlama yapmasıdır (Atılğan, 2005a).

Güvenirliği etkileyen hatalar incelendiğinde, ölçmeyi yapan kişiden kaynaklanan çeşitli puanlama yanlışlıkları aşağıda açıklanmaya çalışılmıştır:

- **Aşırı Hoşgörülü ve Aşırı Katı Olmaktan Kaynaklanan Hatalar:** Puanlayıcının, puanlandığı kişiye yakınlığına göre davranarak bunu puanlamaya yansıtması taraflı ölçmeye neden olmaktadır. Bu bağlamda, aşırı hoş görülü veya aşırı katı olmaktan kaynaklanan bir takım hatalarla birey, hak ettiğinden daha az ya da daha çok puan verilerek değerlendirilmektedir (Morrow ve ark., 2005).

- **Merkezi Eğilim (Standart Ölçme) Hataları:** Bu tür hatalar, puanlayıcıların puanlama sırasında bireylerin performanslarını göz önünde bulundurmaksızın değerlendirme ölçeğinin orta noktasına yakın bir yerlerde puanlamasıdır. Bu durumda puanlayıcılar, bireyleri 5'li bir ölçekte, orta nokta olan 3 aralığında puanlamaktadırlar. Genellikle, bireye yönelik ölçme işlemlerinde ne kadar farklı olursa olsun kişiyi, ortalama veriler çerçevesinde düşünmek ve ortalamaya yakın görmek, hatalı sonuca neden olacaktır (Welk, 2002).

- **En Son Ölçme İşlemine Bağlı Kalma:** Puanlayıcıların, bireylerin

performansını bir önceki değerlendirmeye benzer bir şekilde değerlendirme eğiliminde olmasından kaynaklanan hatalardır (Barutçugil, 2002).

- **Baskın Özellik (Halo Etkisi):** Bu tür hatalar, puanlayıcının, bireyin tek bir pozitif kişilik özelliğini ya da davranışını temel alarak puanlamasından kaynaklanır. Alanyazında halo etkisi olarak da isimlendirilen bu tür hatalar, puanlayıcının, bireyin bir alandaki veya yetideki gelişmiş özelliğinin etkisinde kalarak bunu kişinin tüm özelliklerine genellemesi ile meydana gelir (Welk, 2002).

Eğitim alanında yapılan ölçmeler, bireylerin eğitim sürecinde ele alınan bilişsel, duyuşsal ve devinişsel özelliklere sahip olma düzeylerini belirlemeye yöneliktir. Bu özellikler gizil bir yapıda olduğu için doğrudan gözlenemezler ve bireylerin ölçme araçlarında yer alan maddelere verdikleri gözlenebilir yanıtlar veya performanslar ile dolaylı olarak kestirilmeye çalışılır (Gronlund, 1976).

Geliştirilen performans değerlendirme araçlarının içeriğini oluşturan maddeler hazırlanırken ölçmek istenilen özelliği ne derece ölçtüğüne ve tekrar ölçme sonucunda aynı değerler alıp almadığı hususunda yani ölçme aracının geçerlik ve güvenilirlik düzeyine çok dikkat etmek gerekmektedir. Ölçme aracının geçerlik ve güvenilirlik düzeyinin düşük olması halinde performans değerlendirme sonucu da hatalı olacaktır (Darst ve ark., 1989).

Yukarıda sayılan hataların, puanlayıcı gözlemlerine dayalı ölçmelerin güvenilirliğini olumsuz yönde etkilememesi için aşağıda belirtilen çeşitli önlemler önerilmiştir (Darst ve ark., 1989):

-Puanlayıcılar büyük bir dikkatle seçilmelidir: Puanlayıcıların motivasyonları yüksek olmalı ve en iyisini yapma konusunda istekli olmalıdırlar. İlgi düzeyi düşük puanlayıcılar puanlama sürecini olumsuz etkilerler.

-Puanlayıcılar için etkili bir eğitim programı düzenlenmelidir: Test

sonuçları önemli kararlar vermek için kullanılacaksa, kapsamlı bir eğitim programı düzenlenmelidir.

-Puanlayıcıların değerlendirilmesi: Puanlayıcıların verdikleri puanların yapısı incelenmeli ve verdikleri puanların geçerliliği soruşturulmalıdır.

-İdeal puanlayıcı sayısı: Puanlayıcı sayısının artırılmasıyla, puanlamadaki yanlışlık düşecektir.

-Ölçek maddelerinin ideal sayısı: Ölçek maddelerinin sayısını artırmak puanlayıcının yükünü artırır, ancak olabildiğince çok sayıda ölçek maddesi kullanmak tercih edilmektedir (Darst ve ark., 1989).

Yukarıda belirtilen performansın puanlanmasındaki problemlere ek olarak; puanlama sırasında elde edilen sonuçlar, puanlayıcılar arası anlaşmazlıktan kaynaklanan hatalardan etkilenmektedir. Bu hataların azaltılması amacıyla; puanlayıcıların yeterlik tanımı yapılmalı, puanlayıcıların, puanlama şekilleri oluşturulmuş bir yönerge standardında verilmeli, puanlayıcıların puanlamalarındaki "isteğe bağlılığa" karşı çözümler üretilmelidir (Hambleton ve Plake, 1995).

Araştırma uzun dönemli bir niteliğe sahipse farklı puanlayıcılar arasındaki ölçmeler değişik zamanlarda tekrar edilerek güvenilirliğin devam edip etmediği sınıanmalıdır. Ölçme yapan puanlayıcıların tutarlı puanlar vermeleri için test verme veya ölçme yapma özellikleri konusunda kendilerine video çekimleri izletilmeli veya ölçme işlemi tekrar tekrar yaptırılarak deneyim kazanmaları sağlanmalıdır. Puanlayıcılar arası tutarlılık ancak % 80 düzeyine geldiği zaman kendilerinden bağımsız puanlayıcı olarak yararlanılmalıdır (Şencan, 2005).

Ölçme sonuçlarına karışabilecek hataların kaynaklarına yönelik güvenilirlik hesaplama yöntemleri, ölçme kuramlarının başlıca çalışma alanlarıdır. KTK ve MTK'nın birçok modeli tek bir analizle, tek bir değişkenlik kaynağından gelen hatayı dikkate alır. Oysa birçok ölçme durumunda potansiyel hata kaynağı (maddeler, zaman, puanlayıcılar vs.) ya da değişkenlik kaynağının birden çok

olduğu saptanmıştır.

Sporcunun yetisini ölçmek için puanlayıcıların ölçme sürecine katılmaları gerekmektedir. Puanlayıcıların, ölçme sürecine katıldıkları ölçme uygulamalarında, güvenilirlik çalışmaları için birden çok değişkenlik kaynağının birlikte değerlendirilerek kullanılması gerekmektedir. Bu tür ölçmelerin yapıldığı durumlarda klasik ölçme yaklaşımlarıyla hesaplanan güvenilirlik katsayısı sınırlı kalmaktadır. Bu bölüme kadar açıklanan alanyazındaki kavramların ışığında değişkenlik kaynaklarından gelen potansiyel hataların birlikte değerlendirilmesi ve ortak bir güvenilirlik katsayısının bir analizle hesaplanmasının gerektiği ön plana çıkmaktadır.

Amaç

Futbolcuların yetilerinin tespit edilmesinde kullanılan Futbol Yetilerine İlişkin Dereceleme Ölçeği'nden elde edilen ölçmelerin; Klasik Test Kuramı ve Genellenebilirlik Kuramı'na dayalı olarak belirlenen güvenilirlik katsayılarının saptanması ve birbirleriyle karşılaştırılması araştırmanın amacını oluşturmaktadır.

Bu genel amaç çerçevesinde aşağıdaki sorulara yanıt aranmıştır:

1) Klasik Test Kuramına göre;

a. Futbol Yetilerine İlişkin Dereceleme Ölçeğine ilişkin puanlamalar arası tutarlılığın derecesi nedir?

b. Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin toplam ve alt boyutlarına ait iç tutarlılığın derecesi nedir?

c. Futbol Yetilerine İlişkin Dereceleme Ölçeğine göre yapılan puanlamada puanlayıcılar arası tutarlılığın derecesi nedir?

2) Genellenebilirlik Kuramına göre;

a. Futbol Yetilerine İlişkin Dereceleme Ölçeği'nden elde edilen puanların, puanlayıcı sayılarının 3, 4 ve 5 olduğu ve madde sayılarının bir arttırılıp, bir azaltıldığı (4 boyut için ayrı olacak şekilde; $k=60$; $k=52$) koşullarda, kestirilen G ve Phi katsayılarının derecesi nedir?

b. Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin puanlanmasından elde edilen verilerinin çok değişkenli G çalığıması sonucunda alt boyutların kestirilen varyansları ve toplam varyansı açıklama oranları nedir?

c. Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin alt boyutları için madde sayılarının bir artırılıp, bir azaltılması ve puanlayıcı sayılarının 3, 4 ve 5 olduğu koşullarda kestirilen;

c.1. G katsayılarının derecesi nedir?

c.2. Phi katsayılarının derecesi nedir?

3) Futbol Yetilerine İlişkin Dereceleme Ölçeği'nden elde edilen puanların Genellenebilirlik ve Klasik Test Kuramına dayalı güvenilirlik kestirimleri iç ölçütlere göre değerlendirildiğinde yeterli midir?

Önem

Bu araştırmada, Futbol Yetilerine İlişkin bir Dereceleme Ölçeği 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, hakem/puanlayıcı unsuru da öne çıkmaktadır. Bu bağlamda, hakemlerin/puanlayıcıların psikomotor becerileri puanlamalarındaki güvenilirliğin test edilmesi önem kazanmaktadır. Spor bilimlerinde psikomotor davranışları ölçme ile ilgili alanyazındaki 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 alanyazını için yaygınlaşması da beklenmektedir.

Sayıtlar

Futbolcular, pas, top kontrolü, top sürme ve şut futbol yetilerini uygularken gerçek yeteneklerini yansıtmışlardır. Puanlayıcılar, Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin ikinci puanlamasını ilk puanlamadan bağımsız olarak yapmışlardır.

Sınırlılıklar

Araştırma'nın bulguları, 2007-2008 yılları arasında Ankara ilindeki profesyonel futbol takımlarının alt yapılarındaki yaşları 16-18 arasında olan, antrenman yılları 5 ve üzeri olan 72 futbolcu ile sınırlıdır.

Araştırma, futbol yetilerinden olan pas, top kontrolü, top sürme ve şut yetilerini ölçmeyi amaçlayan 56 maddelik ve 5 seçenekli bir ölçekten elde edilen verilerle sınırlıdır. Araştırmada, G kuramı çalışmalarında, bmxpx desende, belli bir bireyin (b), belli bir maddeden (m) ve puanlayıcıdan (p) elde edilen puanlar dikkate alınarak analizler yapılmıştır.

Tanımlar

Futbol Yetisi: Oyuncunun topla uyumu ve ilişkisi olarak müsabakadan ayrı tek bir hareket performansı şeklinde tanımlanabilir (top sürme, top kontrolü, şut ve pas vb.).

Genellenebilirlik Çalışması: Bir ölçmede, birden çok değişkenlik kaynağını ekonomik ve mantıksal bir esneklik içerisinde değerlendirerek

değişkenleri tahmin eden özel olarak desenlenmiş bir çalışmadır.

Karar Çalışması: G çalışması desenindeki ölçme yöntemini kullanarak özel amaçlar doğrultusunda hatayı en aza indiren çalışmadır.

Değişkenlik Kaynağı (Facet): Potansiyel ölçme hata kaynağı olarak tanımlanan görev, zaman, puanlayıcı gibi ölçme yönteminin bir karakteristiğidir.

Koşul (Condition): Bir değişkenlik kaynağının seviyesi veya düzeyi olarak tanımlanmaktadır.

Kısaltmalar

G Kuramı	: Genellenebilirlik Kuramı
KTK	: Klasik Test Kuramı
G Çalışması	: Genellenebilirlik Çalışması
K Çalışması	: Karar Çalışması
Phi Katsayısı	: Güvenirlilik Katsayısı
b	: Birey temel etkisi
m	: Madde temel etkisi
p	: Puanlayıcı temel etkisi
bxm	: birey ve madde ortak etkisi
bxp	: birey ve puanlayıcı ortak etkisi
m _x p	: madde ve puanlayıcı ortak etkisi
bxm _x p	: birey, madde ve puanlayıcı ortak etkisi

BÖLÜM II

KAVRAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR

Bu bölümde kavramsal çerçeve incelenmiş ve araştırma ile ilgisi kurulabilen bazı çalışmalar ele alınmıştır.

Psikometride Kullanılan Başlıca Test Kuramları

Bilimsel çalışmalarda ortaya atılan ölçme kuramlarının kuramsal yapıdan elde edilen ilişkilerinin pratikte doğruluğunun yoklanması, elde edilen verilerden yola çıkarak ilke ve genellemelere ulaşılması, bu genellemelerin değişik koşullar için doğruluğunun test edilmesi sonucunda kanunlara varılması gereklidir. Aynı zamanda, aynı amaçla ortaya atılmış ölçme kuramlarının aynı koşullar altında karşılaştırılarak, bilgi sağlama güçlerinin, işleyen ve işlemeyen yönlerinin belirlenmesi, pratikte daha kullanışlı, daha doğru olanın saptanması, kuramların benzer ya da farklı yanlarının ortaya konulması bilimin bir işlevi olarak gereklidir (Atılğan, 2004).

Bilim dallarının pek çoğundaki ilerlemeler ölçme araç ya da yöntemlerinin geliştirilmesi ile sağlanabilmiş ve bu gelişme ölçme araç ve yöntemlerindeki gelişmeye paralel olarak hızlanmıştır. Ölçmenin bilimdeki önemi, deneysel yöntemin kurulmasına temel oluşturan güvenilir ve geçerli ölçme sonuçları elde edilmesindedir.

Ölçme çalışmalarında amaç, ölçülen özellik bakımından elde edilen puanın olabildiğince gerçek puana yakın olmasının sağlanmasıdır. Çoğu zaman ölçme sonuçlarına dayalı olarak önemli kararların verilmesi söz konusudur. Kararların isabetliliği yapılan ölçme işleminin 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 için Klasik Test Kuramı (KTK), Madde Tepki Kuramı (MTK) ve Genellenebilirlik kuramı (G) geliştirilmiştir.

Her üç kuram da ölçme araçlarının ve elde edilen ölçme sonuçlarının hatasız olması ile ilgilenmektedir (Baykul, 2000).

I. Klasik Test Kuramı

Klasik Test Kuramı (KTK), eğitim ve psikoloji alanlarında var olan bazı yapıları ölçme yoluyla kuramsal yapılara bağlar ve açıklanmalarına yardımcı olur. Fakat ölçmeye karışan çeşitli hatalar yüzünden ölçmek istediğimiz özelliğin gerçek değeri doğrudan elde edilemez. Bu nedenle ölçme sonucundan elde edilen gözlenen puanlar yardımıyla kestirilmeye çalışılır. Klasik Test Kuramında, ölçülmek istenen özelliğe ait gerçek değer, gerçek puan olarak adlandırılır ve ölçme yoluyla doğrudan elde edilemeyen gerçek puan, bazı varsayımlarla gözlenen puanlardan kestirilir. Bu nedenle Klasik Test Kuramına gerçek puan kuramı' da denir (Baykul, 2000).

Gerçek puanlar doğrudan gözlenemediği için, bazı kuramsal sorunlara yol açar. KTK' ndaki gerçek puan için farklı bakış açıları bulunmaktadır. Öncelikle gerçek puan, doğrudan ölçülemez, kuramsal bir önemi yoktur, "platonik" bir semboldür. Bundan dolayı gerçek puan uygulamada elde edilemediği için, gerçek puanın dışta tutulması gerekmektedir. Böylece gözlenen puan "ölçüt" alınmalı yaklaşımı benimsenmektedir. Gerçek puan kuramsal olarak kullanışlıdır ve uygulamada doğrulanacak sonuçları ortaya çıkarır (Lord ve Novick, 1968).

Gerçek puan, sayılabilir sonsuz çoklukta tekrarlanmış ölçmelerin bir ortalaması olarak kabul edilir. Ölçülen bireylerin değişmeden aynı kaldığı, ardışık ölçmelerin, önceki ölçmelerden etkilenmediği varsayılır. Ölçmelerin tamamen hatadan arınık olması beklenmeyen bir durum olmasına rağmen, KTK ile hatayı en aza indirerek gözlenen puanın gerçek puana yakın bir değer alması olası bir durumdur. Bu amaca ulaşabilmek için KTK aynı özelliğin birden fazla ölçülmesi yolunu kullanır. Çünkü hata, ortalaması sıfır olan tesadüfî bir değişkendir ve ancak sonsuz sayıda ölçme yapıldığı zaman bu değer sıfır olur. Bu sebeple sonsuz sayıda yapılan ölçme sonucunda elde edilen kişinin gözlenen puanının

ortalaması, o kişinin gerçek puanına eşit olur (Hambelton ve Swaminathan, 1985).

Psikometride kullanılan testler de, aynı özelliği ölçmek için sonsuz sayıda uygulama yapılamayacağından dolayı kişinin gerçek puanı tam olarak kestirilemez. KTK'da bireylerin belirli bir psikolojik özelliğe sahip oluş derecelerinin belirlenmesi her bireyin o özelliği ölçmek üzere hazırlanmış olan ölçme aracının maddelerine vermiş oldukları tepkilerin toplamının alınması ile elde edilir. Yani bireyin ölçme aracından almış olduğu ham puan onun ölçülen özelliğe sahip oluş derecesinin bir göstergesidir (Fan, 1998).

KTK'da, ölçmedeki hatanın tesadüfî olduğu ve bu hataya ait puanın, her ölçme için gerçek puandan bağımsız olduğu ve ölçmedeki hata puanı, bir başka ölçmedeki hata puanından bağımsız olduğu kabul edilir. Bu çerçevede KTK'nın temel varsayımları aşağıdaki gibi açıklanabilir:

- Gözlenen puan, gerçek puan ile hata puanının toplamıdır.
- Bir ölçme evreni üzerinde, hata puanları, gerçek puanlardan bağımsızdır. Bir başka deyişle, gerçek puan ve hata puanları arasında ilişki yoktur.
- Her ölçmedeki hata puanları, birbirinden bağımsızdır. Yani, herhangi iki dizi ölçmeye ait hata puanları arasındaki ilişki sıfır'dır.
- Sayılabilir sonsuz çoklukta tekrarlanan ölçmelerdeki hata puanlarının ortalaması sıfır' a yaklaşır (Hambelton ve Swaminathan, 1985).

KTK'nda, ölçme hatalarının bütün yetenek düzeylerindeki bireyler için aynı olduğu varsayılmaktadır. Yani, ölçme hatalarına ilişkin varyansın tüm cevaplayıcılar için aynı kalması, KTK'nın bir sınırlılığıdır. Bazı cevaplayıcılar bir testte diğer cevaplayıcılardan 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 KTK'na dayalı olarak geliştirilen ölçme araçlarının psikolojik özellikleri kestirmedeki gücünün ve test puanlarına ilişkin

verdiği bilginin, yetersiz ve eksik olmasından kaynaklanmaktadır (Hambelton ve Swaminathan, 1985).

KTK'nın bu varsayımlardan kaynaklanan sınırlılıkları incelendiğinde aşağıdaki kavramlardan söz edilmektedir:

- Madde güçlük ve madde ayırıcılık indekslerinin değeri, testin uygulandığı cevaplayıcı grubunun yeteneğine bağlı olmaktadır.

- Testler, orta yetenek düzeyindekiler için geliştirilmiştir. Bu nedenle, düşük ve yüksek yetenek düzeyindeki gruplar için doğru yetenek hesaplamaları elde edilememektedir.

- Güvenirlik kavramının paralel formlarla tanımlanmış olması, güvenilirlik hesaplamalarında, kaynağı belli olmayan tesadüfi hatalardan dolayı, güvenilirliğin alt sınırı elde edilir.

- Ölçme hatalarının varyansı tüm cevaplayıcılar için aynı kalmaktadır.

- Bir testin güçlüğü, bireyin yetenek düzeyi ile eşlendiğinde, belli bir test maddesini doğru cevaplama olasılığının ne olacağı kestirilemez. Bu durum, özellikle, bir veya daha fazla cevaplayıcı grubunda test puanlarının özelliklerini gözlemek isteyen test geliştiricisi için önemlidir.

- KTK, test geliştirme, yanlış maddelerin belirlenmesi, test puanlarının eşitlenmesi, maddelerin maksimum ayırıcılık gücü ve ayırdıkları yetenek ranjı hakkında bilgi vermekte yetersiz kalmaktadır (Hambleton ve Swaminathan, 1985).

Bir ölçme aracındaki herhangi bir maddeye verilen yanıtlar (X), bu maddenin ölçmeye çalıştığı özelliğe öğrencinin sahip olma düzeyi (T) ve hata terimleri (E) ile açıklanır. Bu bağıntı KTK'nın $X=T+E$ modeliyle ifade edilir. Eğer i maddesi için tesadüfi hatalardan arınık bir ölçme yapıldıysa, bu koşulda ölçme aracından elde edilen puan aynı zamanda bireyin ölçülmek istenilen özelliğe sahip olma düzeyine eşit ($X=T$) olacaktır (Fan, 1998).

Uygulamada, sadece gözlem değeri (X) olan ölçme değeri bilinebilir. 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. Bu varsayımdan hareketle ölçmenin varyansı $\sigma^2_x = \sigma^2_t + \sigma^2_e$ şeklinde formüle edilebilir (σ^2_x :Gözlenen değerlerin varyansı, σ^2_t : Gerçek değerlerin varyansı ve σ^2_e : Ölçme hatalarının varyansı) (Gürsakal, 2001).

Ölçme konusu olan bireyler ya da objeler arasındaki farklılıklardan gelen varyans gerçek puanlar varyansı olup, bunun bir birey ya da obje üzerindeki ölçme işlemleri boyunca sabit olduğu varsayımı ile bireyler ya da objeler arası gözlenen varyans; $KO_b = K\sigma^2_{(T)} + \sigma^2_{(E)}$ eşitliği ile ifade edilebilir. Benzer şekilde puanlayıcılara ait varyansa da puanlayıcılar arasındaki farklardan ve hatalardan gelen varyans olmak üzere; $KO_p = N\sigma^2_{(T)} + \sigma^2_{(E)}$ eşitliği yazılabilir. Toplam varyans, $KO_T = [K\sigma^2_{(T)} + \sigma^2_{(E)}] + [N\sigma^2_{(T)} + \sigma^2_{(E)}]$ eşitliği ile ve kalan varyans ise; $KO_k = KO_T - (KO_b + KO_p)$ eşitlikleri ile ifade edilir. KO_k 'nın hata puanları varyansı $[\sigma^2_{(E)}]$ olduğu dikkate alınarak, gerçek puanlar varyansı;

$$\sigma^2_{(T)} = \frac{KO_b - KO_k}{K} \text{ olarak ifade edilir (Lord ve Novick, 1968).}$$

KTK'na göre, $\sigma^2_x = \sigma^2_t + \sigma^2_e$ olduğundan, gözlenen puanların varyansı;

$$\sigma^2_{(X)} = \frac{[KO_b + (K - 1)KO_k]}{K} \text{ olur. KTK, güvenilirliği gerçek puanların varyansının,}$$

gözlenen puanların varyansına oranı olarak tanımlar. Varyans analizi yoluyla yukarıda hem gerçek puanların, hem de gözlenen puanların varyansları tanımlanmış olduğundan, bu iki varyansın oranlanması ile elde edilen

korelasyon, K puanlayıcıdan yalnız birinin verdiği puanların güvenilirliği olur (Baykul 2000).

Tüm puanlayıcıların verdikleri puanların birlikte güvenilirliğinin hesaplanması için; Sperman-Brown formülünde yukarıda belirtilen yolla hesaplanan bir puanlayıcının güvenilirliği yerine konulup, gerekli kısaltmalar ve

düzenlemeler yapılarak, $r_{(T)} = \frac{KO_b - KO_p}{KO_b}$ eşitliği elde edilerek kullanılabilir

(Baykul 2000; Thorndike ve ark., 1991).

KTK'nda güvenilirlik indeksi, gözlenen puanlarla gerçek puanlar arasındaki korelasyonun karesi veya gerçek puanlar varyansının gözlenen puanlar varyansına oranı olarak tanımlanır. Gözlenen puanların varyansı ise, gerçek puanlar varyansı ve hata varyansının toplamından oluşur. Bu durumda

güvenirlik; $P^2(X,T) = \frac{\sigma^2_{(T)} - \sigma^2_{(X)}}{\sigma^2_{(X)}}$ şeklinde ifade edilebilir (Lord ve Novick,

1968; Yurdugül, 2006).

Ölçmele sağlanan bilgilerin kararlı özellik taşıdığına, yani hatadan arındırılmış olduğuna ve aynı amaçla yapılacak ikinci bir ölçmede aynı sonuçların elde edileceğine güven duyulması gerekir (Sencer ve Sencer, 1978).

Ölçeğin taşınması gereken özelliklerden birisi olan güvenilirlik, bir ölçme aracıyla aynı koşullarda tekrarlanan ölçmelerde elde edilen ölçme değerlerinin kararlılığının bir göstergesidir. Güvenirlik, görgül yollarla saptanır ve sayısal bir değerle ifade edilir. Bu sayısal değer genellikle bir korelasyon katsayısıyla ifade edilmesine karşın, güvenilirlik katsayısı 0-1 arasında bir değer alır. Bu özellik güvenilirlik katsayısının istatistiksel tanımından kaynaklanır (Carey,1988; Erkuş, 2003).

Özçelik (1981)'e göre güvenilir olmayan bir ölçek kullanışsızdır. Somut özelliklerin ölçülmesinde kullanılan ölçeklerle yapılan tekrarlı ölçmelerde her seferinde birbirine çok yakın sonuçlar alınır. Soyut özelliklerin ölçülmesinde kullanılan ölçeklerde ise aynı sonuçların alınması çok zordur. Dolayısıyla somut özelliklerin ölçülmesinde kullanılan ölçekler daha güvenilirdir. Birimlerin soyut özelliklerinin ölçülmesinde kullanılan ölçekler diğerlerinden önemli bazı ayrılıklar göstermektedir. Bu tür ölçekler, birimleri ölçmek üzere, belirlenmiş olan standart aralıkları birbirine eşit olsa bile, başlangıç noktaları keyfi olması nedeniyle miktar anlamı olan ölçmeler verememektedir. Bu nedenle, ölçeğin ölçtüğü özelliğin gerçek değerine yakın ölçmeler verme gücünün dikkatle incelenmesine ve dolayısıyla bu incelemede analitik olmayan yani bir ölçme aracıyla elde edilen ölçmelerdeki değişkenliği aşağıdaki gibi ayırmaya gerek vardır;

- i. Ölçülmek istenen özellikteki gerçek farklardan kaynaklanan değişkenlik
- ii. Ölçülmek istenmeyen özelliklerin, ölçülmek istenen özelliği etkilemesi sonucunda gerçek farklardan kaynaklanan değişkenlik,
- iii. i ve ii'nin etkileşimden kaynaklanan değişkenlik,
- iv. Ölçme hatasından kaynaklanan değişkenlik (Özçelik, 1981).

Alanyazında, aynı gözlemlerin paralel testlerin bir setinden daha çoğuna ait olduğu düşünülebileceği, bu nedenle aynı gözlemin birden çok güvenilirlik katsayısına sahip olabileceği bahsedilmektedir. Nitekim çeşitli alt testlerden oluşan bir ölçmede iç tutarlılık güvenilirliği düşük olma eğilimindeyken, test-tekrar test ya da paralel formlar güvenilirliği yüksek olabilmektedir. Bu çelişki ve sınırlılığın temel nedeni; ölçme sonuçlarına karışan hataların Klasik Test Kuramında sadece bir kaynaktan gelen hatalar olarak ele alınmasıdır. Ancak KTK'nın güvenilirlik hesaplama yöntemleri, güvenilirliğin anlamına göre ve ele alınan hata kaynağına göre farklılıklar gösterir (Baykul, 2000; Lord ve Novick, 1968).

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 güvenilirlik belirlemede birbirinin paraleli

olan formlardan gelen hatalar söz konusu olduğundan, formlar hata kaynağıdır. 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).

Devamlılık ya da kararlılık anlamındaki test-tekrar test yönteminde hesaplanan güvenilirlikte hata kaynağı olarak zaman ele alınır. Puanlama tekrar puanlama yönteminde bir testin aynı koşullarda farklı zamanlarda uygulanmasından elde edilen sonuçların benzer olması beklenir. Bu nedenle test-tekrar test yöntemi ile hesaplanan güvenilirlik ölçme aracının zamandan zamana kararlı ölçmeler yapabilme derecesi olduğundan hata kaynağı zaman olarak ele alınır (Crocker ve Algina, 1986).

Bu bölüme kadar KTK'ndaki varsayımlar ve bu varsayımlardan kaynaklanan sınırlılıklardan bahsedilmiştir. Ölçme sonucunda gerçek puana ulaşmak için hatanın kontrol edilmesinin önemi ölçme ile ilgili çalışmalarda ön plana çıkmıştır. Bu bağlamda hata kaynaklarından olan puanlayıcıya ait hata kaynaklarının azaltılması ile ilgili yöntemler aşağıda açıklanmaya çalışılmıştır.

Puanlama Güvenirliğini Kestirme Yöntemleri

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

Bir ölçme aracını geliştirilen kişinin, hem ölçme aracı geliştirme bilgisine, hem de ölçme aracının ölçmeye çalıştığı nitelik ile ilgili alan bilgisine sahip olması gerekir. Bu durum, ölçme aracının güvenilirliğini olumsuz etkileyebilir (Crocker ve Algina, 1986).

Mabry (1999)'e göre, yeteneğin ölçülmesinde kullanılan testlerden elde edilen değerlendirmeler için testi alan kişiden, performansını tekrar sergilemesi istenmez. Çünkü bu performansın yinelenmesi, ikinci performansı etkileyen ve/veya geliştiren bir durumdur. Performans ölçmelerinde güvenilirlik, puanlayıcıların tek bir performansa verdikleri puanlar arasındaki uyum anlamına gelmektedir. Bu puanlar uyumluysa, puanlayıcılar arası güvenilirlik sağlanmış demektir. Puanlayıcılar arası güvenilirlik, aynı performansa farklı puanlayıcılar tarafından verilen puanların tutarlı olması anlamını taşır (Akt: Yıldırım, 1999).

Puanlayıcılar arası uyum ve güvenilirlik, ölçme sonucunda elde edilen puanların iki ya da daha çok puanlayıcı (gözlemci, hakem, yargıcı vb.) tarafından verilmesi koşulunda kullanılır. Puanlayıcılar arası uyum (agreement) ve puanlayıcılar arası güvenilirlik (reliability) arasında fark olduğu Crocker ve Algina (1986) tarafından açıklansa da, çoğu zaman araştırmacılar ve uygulayıcılar arasında aynı anlamda kullanılma eğilimindedirler (Akt: Goodwin, 2001).

Puanlayıcıların verdikleri puanlar arasındaki uyuşmayı belirlemek için öncelikle verilerin niteliği, örneklem büyüklüğü, verilerin büyüklük sırasına konulması ve normal dağılım gösterme durumu araştırılmalıdır (Şencan, 2005).

Puanlayıcılar arası güvenirliliğin hesaplanmasında, puanlayıcıların bireylerin ölçülen özellikleri boyutunda verdikleri puanların ortalaması ve toplamının alınması olanaklıdır. Puanlayıcıların verdikleri puanlar iki boyutlu bir matriste, N bireyin veya objenin her biri üzerinde K puanlayıcıların ölçme sonucu verdikleri puanlar olarak gösterilebilir. Başka bir ifadeyle, K tane puanlayıcının (sütunda) N bireye ya da objeye (satırda) verdikleri puanlar iki boyutlu bir matriste gösterilebilir. Bu şekilde oluşturulacak iki boyutlu bir matristen toplam varyans, satır ortalamalarının varyansı ve sütun ortalamalarının varyansı olmak üzere üç ayrı varyans hesaplanabilir. Satır ortalamalarının varyansı evrende ölçülen özellik bakımından gerçek farkların varyansı ile ölçme hataları varyansı olarak ayrıştırılabilir. Aynı biçimde sütun ortalamalarının varyansı da

puanlayıcılar arasındaki farkların varyansı ve ölçme hataları varyansına ayrılabilir. Puanlayıcılar verdikleri her bir puanı, verdikleri diğer puanlardan bağımsız olarak vermiş ve puanlayıcılar puanlamada diğer puanlayıcılardan etkilenmemişlerse, varyans analiziyle bulunacak olan kalan kareler toplamı, hem puanlayıcılar arasındaki farklılıklardan hem de bireysel farklılıklardan arındırılmış olur (Thorndike ve ark., 1991).

Gözlenen değerlerin standart sapması arttıkça veya güvenilirlik azaldıkça ölçmenin standart hatasında artış olacaktır. Güvenirlik katsayılarının farklı madde yapılarında farklı değer üretmelerinden dolayı, farklı koşullara yönelik olarak hesaplamak için yöntemler geliştirilmiştir (Gürsaka, 2001; Osburn, 2000).

Bu bağlamda, puanlayıcıların puanlamalarına ilişkin güvenirliliğin belirlenmesinde kullanılabilen farklı yaklaşımlar ve teknikler elde edilmiştir; Cohen'in Kappa (1960) ve "Ağırlıklandırılmış Kappa" (Weighted Kappa) istatistikleri, Pearson korelasyon katsayısı, ortalamaların karşılaştırılması (t testi veya varyans analizi) vb. gibi tekniklerin yanı sıra, değişkenlik kaynağı olarak puanlayıcıları ve puanlayıcılardan gelen potansiyel hataları değerlendirerek, puanlayıcılar arası uyum ve puanlayıcılar arası güvenirliliği hesaplamakta G kuramı (Cronbach ve ark., 1972) ve Çok Değişkenlik Kaynaklı Rasch Ölçme (Many-facet Rasch Measurement) modeli de daha yeni sayılabilen yöntemler olarak kullanılmaktadır (Brennan, 2001).

Puanlayıcıların puanlamalarının güvenirliliğini saptamak için kullanılan hesaplama yöntemleri ve istatistiksel analizler aşağıda belirtilmiştir.

Uyuşma İndeksi

Puanlayıcıların uyuştukları madde sayısının toplam değerlendirme veya gözlem sayısına olan oranı olarak tanımlanan uyuşma indeksi, iki puanlayıcının performans, davranış, kompozisyon tipi sınav vb. birçok durumda verdikleri puanların ya da aynı puanlayıcının aynı davranışı iki kez puanlaması durumunda

puanlayıcıların verdikleri puanların uyumunun basit yüzdesi olarak kullanılır. Uyuşma indeksi tekniğinin en belirgin avantajı, hesaplanması ve anlaşılmasının oldukça kolay olması yanında sınıflama, sıralama, aralık ve oran ölçeği değerindeki tüm veriler üzerinde kullanılabilir olmasıdır. Bu yöntemin olumsuz yönü tesadüfî uyuşmaları dikkate almamasıdır. Madde veya görev için uyum yüzdesinin hesaplanması puanlayıcılar arasındaki uyumsuzluk noktalarının tam olarak belirlenmesini sağlayabilir. Böyle bir belirleme puanlayıcıların puanlama davranışlarına ilişkin olarak eğitilmesi için kullanılabilir olması nedeniyle önemli olabilir. Puanlayıcıların tamamen tesadüfî olarak puanlama yapmaları durumunda uyum düzeyi beklendiği gibi çıkmaz. Yani tamamen tesadüfî bir puanlamada uyum, şans kadar ya da tesadüfî değişken oranı kadar beklenirken daha büyük çıkabilir (Goodwin, 2001).

Puanlayıcılar arası değerlendirme sonuçlarının güvenilir sayılabilmesi için Uyuşma indeksi değerinin %75'in üzerinde olması gerekir. Daha düşük bir oran, değerlendirmede puanlayıcıların önemli ölçüde farklı düşündükleri anlamına gelir. Uyuşma indeksi = Toplam Uyuşma Sayısı / Toplam Değerlendirme Sayısı*100 hesaplamasıyla elde edilmektedir (Şencan, 2005).

Cohen Kappa Katsayısı

Cohen kappa katsayısı, daha az sınırlayıcı olan ve daha geniş bir uygulama alanı bulmaktadır. Cohen'in kendi adıyla birlikte anılan Cohen'in Kappa (*K*) katsayısı şans düzeyini dikkate alarak uyum düzeyini kestirmeyi sağlar. Puanlayıcıların verdikleri puanlar aynı ölçme içinde sadece beklenen şans düzeyi kadar uyum gösterirlerse *K* sifıra eşit olur. Uyum düzeyinin beklenen şans düzeyini aşması durumunda ise *K* sıfırdan büyük olur (Goodwin, 2001).

Kappa istatistiğinin sınıflama verileriyle kullanılarak bir uyum ölçüsü elde edilmesinde kullanılabileceği farklı desenler vardır. *K* kategori sayısı ve *M* puanlayıcı sayısı olmak üzere, sınıflama verileriyle en basit ve en yaygın güvenilirlik kestirimi, iki puanlayıcı (*M*=2) ve iki kategori (*K*=2) olması (2x2)

durumudur. Kappa istatistiđi iki veya daha fazla puanlayıcının yaptıđı deđerlendirmeler arasındaki uyuşmayı belirlemek için kullanılır. Kappa istatistiđinin uyuşma indeksinden farkı, şans faktörünün etkisini ortadan kaldırmasıdır. Kappa katsayısı -1 ile +1 arasında deđişir. Sıfır deđerı tesadüfi uyuşmayı, negatif deđerler tesadüfi olmaktan daha kötü bir uyuşmayı, +1 ise yüksek düzeyde uyuşmayı temsil eder. Kappa katsayısı 0,40 ile 0,75 arasında ise makul bir uyuşma, 0,75'ten büyük ise mükemmel bir uyuşma olduđu anlamına gelir. Uyuşma oranı çok yüksekse, deđerlendirmelerin dođru olmama ihtimali de göz önünde bulundurulmalıdır (Helms, 1999).

Kendall Uyuşma Katsayısı

Kendall uyuşma katsayısı (W), çok sayıda puanlayıcının vermiş olduđu puanlar arasında ne ölçüde uyuşma bulunduđunu belirler. Hakemlerin veya puanlayıcıların verdikleri puanlar sıralı ölçek verisi niteliğinde olmaktadır (Şencan, 2005).

Martin ve Bateson (1992)'e göre; farklı puanlayıcılar tarafından gözlenen davranışlar için puanlayıcılar arasındaki tutarlılıđı saptamak amacıyla, puanlayıcılar arası güvenilirliđin saptanması gerekmektedir. İkiden fazla puanlayıcı varsa güvenilirlik için hesaplanması gereken korelasyon katsayısı, parametrik olmayan istatistiksel bir teknik olan Kendall'ın Uyuşum katsayısıdır (Akt: Tavşancıl, 2002).

Howell (2002)'e göre puanlayıcılardan alınan ađırlıklı seęenek puanları arasında uyum olup olmadıđını test eden Kendall'ın W uyuşum katsayısı, n nesnenin veya kişinin ikiden fazla sıralanışı arasındaki bađıntının derecesini veren bir korelasyon ölçüsüdür. Bu ölçü, özellikle yargılar arası veya testler arası güvenilirlik araştırmalarında kullanılmakta, sıralanabilir veri için puanlayıcılar arası güvenilirlik indeksi olarak tanımlanmaktadır (Akt: Çıtak, 2007).

İkiden daha fazla deđerşken söz konusu olduđunda puanlayıcılar arası

korelasyonu ölçmek için kullanılan W testi; k gözlem ya da sıraların kümesi arasındaki uyumu ölçmeye yöneliktir. İlk yaklaşımda k tane gözlem, b tane özelliğin her birinde elde edilen puanlara göre sıralanabilir. İkinci yaklaşımda ise b adet karar verici k tane gözlemi aynı özelliğe dayalı olarak sıralar. Bu tür durumlarda kullanılan Kendall W uyum katsayılarının varsayımları aşağıdaki gibidir.

- Veriler k gözlem ya da birey üzerinde b adet ölçmeden oluşur.
- Ölçme ölçeği en azından sıralıdır.
- Kaydedilen gözlemler sıra sayılarından oluşabilir. Orijinal veriler sıralı değilse sıralanmalıdır (Kıroğlu, 2001).

Kendall'in uyum katsayısı kullanılarak yapılan değerlendirmelerde puanlayıcılar arasındaki uyumun en az 0,80 düzeyinde olması istenmektedir (Szymanski ve Linkowski, 1993).

Pearson Çarpım- Moment Korelasyon Katsayısı

Puanlayıcılar arası güvenilirliğin belirlenmesi için Pearson çarpım-moment korelasyon katsayısının hesaplanması, iki puanlayıcının bireylere, objelere vs. ölçülen özellik boyutunda verdikleri puanların tutarlılığı olarak tanımlanır. Güvenirlik hesaplamada, güvenilirlik katsayısı olarak kullanılan bu korelasyon katsayısı iki puanlayıcının puanlarının doğrusal ilişkisini gösterir (Balcı, 2000).

Pearson çarpım-moment korelasyon analizi; iki puanlayıcının yaptığı değerlendirmeler eşit aralıklı ölçek verilerine dayanıyorsa, verilerin büyüklük sırası önemli değilse ve veri sayısı 30'un üzerinde ise ilişkiyi belirlemek amacıyla kullanılır. Ancak Pearson çarpım-moment korelasyon analizi iki puan dizisi arasındaki ilişkiyi belirler. Puanlayıcıların yüzde kaç oranında uyuştuklarını göstermez. Puanlayıcılar arasındaki varyansı dikkate almadığından puanlayıcılar arasındaki değişkenliğe duyarsızdır (Şencan, 2005).

Hartman (1977)'e göre Pearson korelasyon katsayısının hesaplandığı

grubun büyüklüğünden doğrudan doğruya etkilenmesi bir dezavantajdır (Akt: Goodwin 2001). Puanlayıcılar arası güvenilirliğin hesaplanmasında Pearson korelasyonunun kullanılması daha önemli bir dezavantajı ise, ortalamadan bağımsız olmasıdır. Bu nedenle, güvenilirlik hesaplamada iki puanlayıcının puanlama düzeyleri arasındaki benzerlik ve farklılıklarını gösteremez. Başka bir ifadeyle, Pearson çarpım-moment korelasyon katsayısı iki puanlayıcının verdikleri puanların doğrusal ilişkisi yani bu iki değişkenin birlikte değişiminin bir ölçüsüdür (Baykul, 2000).

Pearson korelasyon katsayısının puanlayıcılar arası uyumun ya da güvenilirliğin belirlenmesindeki dezavantajını ortadan kaldırmak için varyans analizine dayalı olarak gruplar-içi korelasyonu hesaplanmalıdır. Varyans analizi yoluyla hesaplanabilen bu güvenilirlik katsayısı 0 ile + 1 arasında değerler alır. Bu katsayı varyans analizi literatüründe grup-içi korelasyon (Intraclass Corelation Coefficient) denilen özel bir korelasyon katsayısıdır (Goodwin, 2001).

Puanlayıcılar, yapılan ölçümlerde benzer puanları vermişlerse sonuçlar güvenilir demektir. Ölçme aracı kullanılarak yapılan değerlendirmelerde puanlayıcılar arasındaki uyuşmanın en az 0,80 düzeyinde olması istenir. Ölçme sadece gözlemlerde bulunmak amacıyla yapılmışsa puanlayıcılar arasındaki uyuşmanın 0,70 olmasının yeterli olacağı belirtilmiştir (Szymanski ve Linkowski, 1993).

KTK'nda puanlayıcılar arasındaki uyuşmanın bulunması amacıyla uygulanan farklı yöntemler puanlayıcı sayısına göre farklılık göstermektedir. Bu yöntemlerden elde edilen katsayılar ile puanlayıcıların puanlamalarındaki tutarlılık yorumlanmaktadır.

II. Madde Tepki Kuramı

Madde Tepki Kuramı (MTK) veya Örtük Özellikler Kuramı, örneklemden bağımsız madde ölçeklemesini ve maddeden bağımsız olarak yetenek kestirimini

matematiksel modellerle olanaklı hale getirmek iddiasındadır. Kuramın temeli, bireylerin yeteneklerini gösteren psikolojik uzayla, bu uzayı temsil etmek amacıyla hazırlanan maddeler arasındaki bağıntılardır. Kurama göre, kişilerin özellikleri bir q uzayından oluşur ve çalışılan özelliğin ölçülmesi amacıyla hazırlanan maddelerden oluşturulmuş testin ise q uzayını temsil ettiğini kabul eder (Hambleton ve Swaminathan, 1985).

MTK, bireyin davranışını, “tretyler” (traits) adı verilen psikolojik özellikleri tanımlama yoluyla, belli bir dereceye kadar açıklanabileceğini, bu psikolojik özelliklerin her birinde, bireyin yerinin puan olarak kestirilebileceğini ve sonra da, benzer koşullarda bireyin davranışını açıklamak ve yordamak üzere elde edilen sayısal değerlerin kullanabileceğini farz eder (Lord ve Novick, 1968).

Hambleton (1989)’a göre; MTK, bireyin testle ölçüldüğü düşünülen gözlenmeyen psikolojik özellikleri (yetenek) ile bireyin o testte gösterdiği cevaplama edimi arasındaki doğrusal ilişkiyi ortaya koymaya çalışır. Gustafsson (1988)’a göre ise MTK, bir cevaplayıcının bir maddeye doğru cevap verme olasılığını gösterecek olan madde özellikleriyle, cevaplayıcının yetenek düzeyinin nasıl birlikte ele alınacağını öngören matematiksel eşitlikler sunarak, iki özellik arasındaki ilişkiyi, KTK’ndan farklı olarak, matematiksel biçimde işe vuruklaştırır (Akt: Demirtaşlı, 1995).

MTK; bir parametrelili, iki parametrelili ve üç parametrelili model adında üç farklı model öne sürmektedir. Bir parametrelili model madde güçlüğü parametresi (b_i) ve bireyin yetenek düzeyi arasında ilişki kurmaktadır. İki parametrelili modelde madde güçlük parametresinin yanına belirli bir yetenek düzeyinde sorunun düşük ve yüksek yetenek grubundakileri ayırma gücü olarak açıklanan ayırt edicilik parametresi (a_i) eklenmektedir. Üçüncü modelde ise iki parametrelili modele düşük yetenek düzeyindeki bireyleri maddeyi doğru yanıtlama olasılığı (c_i) olarak açıklanan şans faktörü parametresi eklenmektedir. MTK’nın zayıf

yanları arasında büyük örneklem gruplarına ihtiyaç duyması, kuramın karmaşık olması, sonuçların yorumlanmasının zor olması ve özel yazılımlar gerektirmesi, KTK'na göre sağlanması daha zor olan bir takım varsayımlar gerektirmesi sayılabilir (Hambleton ve Swaminathan, 1985).

III. Genellenebilirlik Kuramı

KTK'nda ölçme sonuçlarına karışan hatalar, sadece bir değişkenlik kaynağından gelen hatalar olarak ele alınmaktadır. Dolayısıyla aynı ölçmeye ilişkin olarak KTK'nın farklı hata kaynaklarını dikkate alan güvenilirlik yöntemleri ile elde edilen katsayılar da birbirinden farklı olabilir. Puanlayıcıların, bireyleri ilgilenilen özelliklerini derecelendirme durumunda KTK'nın yöntemleri doğrudan uygulanamaz. Bu gibi durumlarda değişim kaynaklarını belirlemek için varyans analizi teknikleri kullanılır. Genellenebilirlik (G) kuramı, davranış ölçmede güvenilirliğin değerlendirilmesini, güvenilir gözlemlerin tasarlanmasını, araştırılmasını ve kavramlaştırılmasını sağlayan istatistiksel bir kuramdır ve varyans analizine (ANOVA) dayalıdır (Shavelson ve Webb, 1991).

KTK çerçevesinde bahsedilen güvenilirlik, üzerinde ölçme yapılan gruba bağımlıdır. Ölçme yapılan grup yani örneklem değiştiğinde, güvenilirlik katsayısının değeri de değişmektedir. G Kuramı, KTK'nı yeniden ele alarak gözlemler evreninden rastgele alınmış örneklemelerin, sadece bir tane rastgele alınmış gözlem örneklenmesi olarak genellenebilme yapmanın elverişsizliğinden dolayı bu kurama karşı geliştirilmiştir. Çünkü gözlemler birçok değişik evrenden sezgisel olarak genellenebilir. Gözlemler, bu evrenler hakkında güvenilir genellenebilmeler verir, birçok değişkene sahip olabilir ve böylece birçok değişik güvenilirlik katsayısıyla da birleştirilebilir. G kuramı ise bireyin evren puanını genelleme evrenindeki ölçmelerin ortalaması olarak tanımlanmasıyla güvenilirliğe değişik bir bakış açısı getirir ve bu ölçmelerin kesin olarak paralel olması gerekmez (Atılğan ve Tezbaşaran, 2005; Brennan, 2001).

G kuramı, günümüzde hala yaygın kullanılan KTK'nın gerçek puan modelinin sınırlılıklarına olan tepkilerden hareketle Cronbach ve arkadaşları

(1972) tarafından ortaya atılmıştır. G kuramıyla yeni bir güvenilirlik formülü geliştirilememesine karşın, güvenilirlik kavramına yeni bir mantık getirilmesi açısından önemli bir aşama oluşturulmuştur (Shavelson ve Webb, 1991).

Puanlayıcıların görev/madde puanlamalarından kaynaklanan hata kaynaklarının çok oluşundan dolayı, tek bir değişkenlik kaynağı ile ilgili güvenirlüğün tanımlanması sınırlı kalmaktadır. Birçok hata kaynağını birlikte tahmin ettiğinden dolayı G Kuramının uygulanması önemli olmaktadır (Baykul, 2000).

G kuramı, sadece KTK'nın güvenilirlik kavramını yeniden yorumlamaz. Aynı zamanda geleneksel olan güvenilirlik ve geçerlik kavramlarının ayrımının yapılmasını da gösterir. Geçerlik türleri içinde yapı geçerliği, gözlenebilen ölçmeler üzerine temellenmiş örtük bir yapı hakkında vardama yapılmasıyla ilgilidir. Geçerliği oluşturma, gözlemlenen ölçmelerin gizli kalmış yapılarına dayanarak alınan çıkarımlar olarak düşünülür. Öyleyse güvenilirlik de, paralel ölçmeler boyunca gözlemlerden alınan gerçek puanlardan elde edilen çıkarımlar olarak düşünülmektedir (Shavelson ve Webb 1991).

KTK'nın sağlanması güç olan paralel testler için eşit ortalama, varyans ve kovaryans varsayımına karşılık, G Kuramının sadece aynı evrenden tesadüfi olarak çekilmiş paralel testlerin örnekleme gibi kolay sağlanabilecek bir varsayımı vardır (Brennan 2001; Shavelson ve Webb 1991).

Alt testlerden oluşturulmuş bir testin güvenirlüğünün hesaplanmasının gerekli olduğu çok değişkenli ölçme durumları için G kuramı genişletilerek çok değişkenli genellenebilirlik (multivariate generalizability) adı ile kullanılabilir. Tek değişkenli genellenebilirlik, tek bir testin uygulanmasından elde edilen puanların analizi için kullanılırken, bir testin farklı alt testlerinden elde edilen puanların genellenebilirliğinin aynı zamanda belirlenmesinde çok değişkenli genellenebilirlik hesaplanabilir. Bir testin alt testlerden oluşması durumunda, eğer sadece puan güvenirlüğünün etkisi ile ilgileniliyorsa, farklı maddelerin bütün testin içindeki

varyans bileşenlerinin kestirilmesi ve puanlarının güvenilirliğinin tek değişkenli genellenebilirlikle hesaplanabilmesi olanaklıdır (Yin ve Shavelson, 2004).

G kuramı; madde, zaman, puanlayıcı ve benzeri hata kaynaklarını, değişkenlik kaynağı olarak adlandırır. Değişkenlik kaynağı, deneysel desenlerin literatüründeki “faktör” kavramına benzer. Bu değişkenlik kaynakların düzeyleri (levels) (madde sayısı, puanlayıcı sayısı), koşullar (conditions) olarak adlandırılır. Genellenebilirlik kuramında kullanılan değişkenlik kaynağı ve koşul ifadeleri geleneksel varyans analizinde faktör ve düzey kavramlarına karşılık gelir. Gözlemlerin evreni, değişkenlik kaynaklarının bütün olası birleşim düzeyleri olarak tanımlanır (Brennan, 2001).

G kuramıyla ölçme hatalarının farklı değişkenlik kaynaklarından kestirilmesi, gruplar arası korelasyon katsayısının hesaplanmasında olduğu gibi varyans analizi yaklaşımının genişletilmiş bir halidir. Bu bağlamda; G kuramının amacı, ölçme konusu olan bireyler ya da objelerin gözlenen puanlarının evren puanlarına (KTK’ndaki gerçek puan kavramına karşılık gelir) doğrulukla genellenmesini sağlamaktır (Atılğan, 2008).

Genellenebilirlik (G) çalışması

G kuramında, genellenebilirlik (generalizability) katsayısı olarak adlandırılan bir katsayı hesaplanır. Bu katsayı KTK’ndaki güvenilirlik katsayısına benzemekle birlikte güvenilirlik kavramını yeniden yorumlamaz. G kuramı, güvenilirlik ve geçerlik arasındaki geleneksel farkın güvenilir gözlemler düzenlemekle nasıl ortadan kaldırılabilceğini de gösterir. G kuramında bir evren, onun değişkenlik kaynakları ve gözlemlerin koşulları geçerlik kavramının geleneksel alanında açıklanan yapıdan tanımlanır. G kuramı, gözlemlerin örtülü yapısı (yani kabul edilebilir gözlemlerin evreni) hakkındaki vardamaları doğrulukla göstermeyi sağlarsa, gözlemleri güvenilir olarak tanımlar (Shavelson ve Webb, 1991).

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

KTK'nda $X = T + E$ eşitliğinden hareketle güvenilirlik gerçek puanların varyansının gözlenen puanlarının varyansına oranı olarak tanımlanmaktadır ve

elde edilen $P(X, X^-) = \frac{\sigma^2(T)}{\sigma^2(X)}$ eşitlikte yer alan gerçek puanların (T) ve hata

puanlarının (E) varyansı hesaplanabilirse güvenilirlik katsayısı elde edilebilir. Bir kişinin evren puanı (Φ) genellemenin evreni içinde bütün gözlemlerden bireyin beklenen puanı olarak tanımlanır (Shavelson ve Webb, 1991).

G kuramı içinde tanımlanan hata puanları varyansı ise, farklı 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 puanların varyansına katkısı ve faktörlerin gözlenen puanlar üzerindeki etkisini araştırır. Bununla birlikte G çalışması eldeki ölçme sonuçlarının var olan durumuna göre bir güvenilirlik kestirir (Brennan, 2001; Shavelson ve Webb, 1991).

G kuramındaki değişkenlik kaynakları içerisinde bulunan tesadüfi değişkenlik, ölçme işlemlerinin koşullarının tesadüfi olarak örneklenmesi anlamına gelmektedir. Tesadüfi değişkenlik, geniş bir puanlayıcı evreninden tesadüfi olarak çekilen bir puanlayıcı örnekleminin puanlama yapmasıdır. Bunun

yanında genellenebilmenin evreninin sonsuz büyüklükte ve değiştirilebilir özellikte olduğu varsayıldığında, ölçme işlemlerinin koşulları tesadüfi olarak çekilmemişse bile, değişkenlik kaynağı tesadüfi olarak kabul edilir. Örneklemin evrenden küçük olduğu ve örneklemin evrenden tesadüfi olarak çekildiği ya da değiştirilebilir olduğu durumlarda değişkenlik kaynağı tesadüfi değişkenlik kaynağı olarak adlandırılır (Yin ve Shavelson, 2004).

Diğer bir değişkenlik kaynağı da sabit değişkenlik kaynağıdır. Sabit değişkenlik kaynağı, özellikle seçilmiş olan ve bunun dışında genellenmesiyle ilgilenilmeyen veya çalışılan evrenin küçük olması nedeniyle evren üzerinde çalışılması durumlarında kullanılır (Brennan, 2001; Crocker ve Algina, 1986).

G kuramında; değişkenlik kaynağının sayısına bağlı olarak desenin oluşturulmasında tesadüfi (random) ve sabit (fixed) olmasının yanı sıra, veri yapısına bağlı olarak çaprazlanmış (crossed) ya da yuvalanmış (nested) olmak üzere iki tür desen bulunmaktadır. Ölçmedeki değişkenlik kaynağının bütün koşulları diğer bir değişkenlik kaynağın bütün koşullarını gözlüyorsa çaprazlanmış olarak adlandırılır ve değişkenlik kaynakların arasına "x" işareti konularak gösterilir. Bir değişkenlik kaynağın bütün koşulları diğer bir değişkenlik kaynağın bütün koşulları tarafından gözlemlenmiyor, sadece bazı koşulları diğer bir değişkenlik kaynağın bazı koşullarınca gözlemleniyorsa yuvalanmış olarak adlandırılır ve iki değişkenlik kaynağı arasına ":" işareti konularak gösterilir. Değişkenlik kaynaklarının sayısına göre; bir, iki ve daha fazla değişkenlik kaynaklı (facets) desenler tanımlanır (Shavelson ve Webb, 1991; Yin ve Shavelson, 2004).

Bireylerin performanslarının puanlayıcılar tarafından puanlanması söz konusu olduğunda, her puanlayıcı her bireyi gözlemleyip puanlıyorsa diğer bir deyişle verilerin çaprazlanmış olması durumunda, birey b, madde m ve puanlayıcı p ile gösterildiğinde simgesel olarak bxm_p şeklinde ifade edilir. Benzer şekilde bireylerin her birinin performansı farklı puanlayıcılar tarafından

gözlemleniyorsa $m \times (b:p)$ ile şeklinde ifade edilir (Atılğan, 2004).

Bu iki farklı desende ölçmenin hedefi bireyler olduğundan b bir hata kaynağı değildir. Bu desenlerde bireylerin puanlarını etkileyen iki bağımsız değişken (tesadüfî faktör-facet) olduğu görülmektedir. G Kuramının $b \times m \times p$ deseninde tanımlanmış geniş bir evrende yer alan bütün test maddeleri ve puanlayıcıların genellenmesinde, belli bir bireyin (b), belli bir maddeden (m) ve puanlayıcıdan (p) elde edilen puanları Tablo 1'de açıklanmıştır (Brennan, 2001).

Tablo 1. G Kuramının $b \times m \times p$ Deseninden Elde Edilen Puanlara Ait Eşitlikler

$\mu_b - \mu$	Birey etkisi
$\mu_m - \mu$	Madde Etkisi
$\mu_p - \mu$	Puanlayıcı Etkisi
$\mu_{bm} - \mu_b - \mu_m + \mu$	Birey x Madde Etkisi
$\mu_{bp} - \mu_b - \mu_p + \mu$	Birey x Puanlayıcı Etkisi
$\mu_{mp} - \mu_m - \mu_p + \mu$	Madde x Puanlayıcı Etkisi
$X_{bmp} - \mu_b - \mu_m - \mu_p + \mu_{bm} - \mu_{bp} + \mu_{mp} - \mu$	Kalan Etki

Tablo 1'deki eşitliklerde tanımlandığı gibi, E beklenen değer olmak üzere $\mu = E_b E_m E_p X_{bmp}$ ve $\mu_b = E_p E_m X_{bmp}$ olarak yazılabilir. Birey etkisinin varyansı $(\sigma_b^2 = E_b (\mu_b - \mu)^2)$ KTK'ndaki gerçek puan varyansına benzer olarak evren puanı $(\sigma_{(\tau)}^2)$ varyansı olarak adlandırılır. Gözlenen puanların varyansı; $\sigma^2(X_{bmp}) = \sigma_b^2 + \sigma_m^2 + \sigma_p^2 + \sigma_{bm}^2 + \sigma_{bp}^2 + \sigma_{mp}^2 + \sigma_{bmp,e}^2$ eşitliği ile tanımlanır (Shavelson ve Webb, 1991).

Gözlenen puanlar varyansındaki her bir varyans bileşeni varyans analizi kullanılarak Tablo 2'deki eşitlikler yardımıyla kestirilmektedir (Brennan 2001).

Tablo 2. İki Değişkenlik Kaynaklı Tesadüfi Desen için Varyans Bileşenlerinin Kestirilmesine Ait Eşitlikler

Varyansın Kaynağı	Kareler Toplamı	Serbestlik Derecesi	Kareler Ortalaması	Kestirilen Varyans Bileşeni
Birey (b)	KT_b	$n_b - 1$	$KO_b = \frac{KT_b}{sd_b}$	$\sigma_b^2 = \frac{KO_b - \sigma_{bmp,e}^2 n_p \sigma_{bm}^2 - n_m \sigma_{bp}^2}{n_m n_p}$
Puanlayıcı (p)	KT_p	$n_p - 1$	$KO_p = \frac{KT_p}{sd_p}$	$\sigma_p^2 = \frac{KO_p - \sigma_{bmp,e}^2 n_b \sigma_{mp}^2 - n_m \sigma_{bp}^2}{n_m n_p}$
Madde (m)	KT_m	$n_m - 1$	$KO_m = \frac{KT_m}{sd_m}$	$\sigma_m^2 = \frac{KO_m - \sigma_{bmp,e}^2 n_p \sigma_{bm}^2 - n_b \sigma_{mp}^2}{n_m n_p}$
bxp	KT_{bp}	$(n_b - 1)(n_p - 1)$	$KO_{bp} = \frac{KT_{bp}}{sd_{bp}}$	$\sigma_{bp}^2 = \frac{KO_{bp} - \sigma_{bmp,e}^2}{n_m}$
bxm	KT_{bm}	$(n_b - 1)(n_m - 1)$	$KO_{bm} = \frac{KT_{bm}}{sd_{bm}}$	$\sigma_{bm}^2 = \frac{KO_{bm} - \sigma_{bmp,e}^2}{n_p}$
pxm	KT_{pm}	$(n_p - 1)(n_m - 1)$	$KO_{mp} = \frac{KT_{mp}}{sd_{mp}}$	$\sigma_{mp}^2 = \frac{KO_{mp} - \sigma_{bmp,e}^2}{n_m}$
bxm _{p,e}	$KT_{bpm,e}$	$(n_b - 1)(n_m - 1)(n_p - 1)$	$KO_{bpm,e} = \frac{KT_{bpm,e}}{sd_{bpm,e}}$	$\sigma_{bpm,e}^2 = KO_{bpm,e}$

Varyans bileşenlerinin göreceli büyüklükleri, davranış ölçmesini etkileyen potansiyel hata kaynakları hakkında bilgi sağlamaktadır. Gruplar arası korelasyon katsayısında olduğu gibi varyans bileşenlerinden F değerleri hesaplanarak istatistiksel olarak anlamlı olup olmadıkları test edilemez. Fakat kestirilen varyans bileşenlerinin standart hataları kestirilen varyans bileşenlerinin örneklem değişkenlikleri hakkında bilgi verir (Brennan, 2001; Shavelson ve Webb, 1991).

G kuramında, Genellenebilirlik çalışmasından farklı olarak Karar (K) çalışması olarak adlandırılan diğer bir çalışma vardır. K çalışmasında, G çalışmasından elde edilen bilgileri kullanarak belli bir amaçla yapılan bir ölçmedeki hataları en aza indirmenin yolları aranır. G çalışmasında yapılan bir ölçmenin güvenilirliği değerlendirilirken, karar çalışması Genellenebilirlik çalışmasından hareketle karar üzerinde temellenen verileri toplamak üzere düzenlenir (Crocker ve Algina, 1986).

Karar (K) Çalışmaları

Genellenebilirlik kuramı, göreceli veya kesin kararlar için, güvenilirliği etkileyen tüm hata kaynaklarını hesaba katar. G kuramındaki Karar (K) çalışmalarıyla (Decision Study) elde edilecek orijinal puanlayıcı-madde sayıları ile farklı puanlayıcı-madde sayıları senaryolarının kestirilmesi yapılabilmektedir. Ölçeğin orijinal hali ile farklı senaryolardan elde edilen sonuçların benzerlik göstermesi ölçeğin kullanılmasının uygun olabileceği şeklinde yorumlanabilecektir (Dimitrov, 2002).

G kuramı, Karar çalışmasını G çalışmasından ayırır. K 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 Algina 1986; Shavelson ve Webb 1991).

G Kuramında, istenilen düzeyde bir güvenilirlik katsayısına ulaşılabilmesi için, bir ölçme durumunun değişkenlik kaynaklarının varyanslarından hareketle, değişkenlik kaynaklarının (facet) koşullarının (condition) sayısal miktarı belirlenebilmektedir. Bu amaçla, G Kuramındaki Karar çalışması, varyanslardan elde edilen bilgileri kullanarak belli bir amaçla yapılan bir ölçmedeki hataları en aza indirmenin yollarını araştırmak için düzenlenir. Bir Karar çalışmasında; genellenenin evreni olarak adlandırılan ve genelleme yapılmak istenilen evrende kabul edilebilir gözlemlerin, evrenin kapsamında yer alan bazı ya da bütün değişkenlik kaynakları ve bu değişkenlik kaynaklarının düzeyleri tanımlanır

(Brennan, 2001; Crocker ve Algina, 1986).

Başka bir ifadeyle, Karar üzerine temellenen verileri toplamak için düzenlenen K 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).

G 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). Bağıl bir değerlendirme, ölçme sonucunda elde edilen puanların gruba bağılı olarak çıkarılmış olan bir norma göre yapılırken, mutlak değerlendirme ise gruba bağılı olmadan önceden belirlenmiş mutlak bir ölçüte göre yapılır. Bu nedenle de G kuramında ölçme hataları bağıl kararlar için ölçme hataları ve mutlak kararlar için ölçme hataları olarak değerlendirilir. Bir başka deyişle; G Kuramında güvenilirlik katsayısı olarak hesaplanan Genellenebilirlik ve Phi katsayılarının hesaplanmasında yer alan bağıl (G-katsayısı için) ve mutlak (Phi-katsayısı için) hata varyanslarını azaltabilmek için alternatif K çalışmaları düzenlenebilir (Brennan, 2001; Shavelson ve Webb,1991).

Genellenebilirlik kuramıyla hesaplanan G (genellenebilirlik) katsayısı bağıl ve mutlak değerlendirmeler için ayrı ayrı hesaplanabilir. İki değişkenli analiz için b_{xm}x_p deseninde yedi adet ana ve birbiriyle etkileşimde olan varyans bileşeni " $\sigma^2_b + \sigma^2_m + \sigma^2_p + \sigma^2_{bm} + \sigma^2_{bp} + \sigma^2_{mp} + \sigma^2_{bmp,e}$ " bir tek G çalışması ile kestirebilir (Brennan, 2001).

K çalışmasında genellenebilirlik (generalizability) katsayısı (G ya da E_p^2) ve güvenilirlik (dependability) katsayısı (Φ (Phi)) olmak üzere iki farklı güvenilirlik katsayısı farklı ölçme senaryoları için hesaplanabilir (Brennan, 2001; Shavelson ve Webb,1991).

Bağıl hata terimi “ δ ” ile gösterilmek üzere bağıl hata varyansı;

$$\sigma^2\delta = \frac{\sigma^2_{bm}}{n_m} + \frac{\sigma^2_{bp}}{n_p} + \frac{\sigma^2_{bmp}}{n_m n_p} \quad \text{eşitliği ile gösterilir. } \sigma^2\delta \text{ bireylerin}$$

puanlarının evren değerinin varyansı olarak KTK'ndaki gerçek puanların varyansına karşılık gelmektedir.

Böylece, genellenebilirlik katsayısı;

$$G = \frac{\sigma^2_b}{\sigma^2_b + \sigma^2\delta} \quad \text{bağıl hata terimi } (\sigma^2\delta) \text{ yerine konularak aşağıdaki bağıl}$$

hata terimi için G katsayısı elde edilir.

$$G = \frac{\sigma^2_b}{\sigma^2_b + \frac{\sigma^2_{bm}}{n_m} + \frac{\sigma^2_{bp}}{n_p} + \frac{\sigma^2_{bmp}}{n_m n_p}}$$

Mutlak hata terimi (Δ) varyansı aşağıdaki eşitlik ile gösterilir;

$$\sigma^2\Delta = \frac{\sigma^2_m}{n_m} + \frac{\sigma^2_p}{n_p} + \frac{\sigma^2_{bm}}{n_m} + \frac{\sigma^2_{bp}}{n_p} + \frac{\sigma^2_{mp}}{n_m n_p} + \frac{\sigma^2_{bmp}}{n_m n_p} \quad \text{Mutlak hata varyansının}$$

yerine konularak Güvenirlik indeksi olarak adlandırılan Φ katsayısı aşağıdaki gibi elde edilir.

$$\Phi = \frac{\sigma^2_b}{\sigma^2_b + \frac{\sigma^2_m}{n_m} + \frac{\sigma^2_p}{n_p} + \frac{\sigma^2_{bm}}{n_m} + \frac{\sigma^2_{bp}}{n_p} + \frac{\sigma^2_{mp}}{n_m n_p} + \frac{\sigma^2_{bmp}}{n_m n_p}}$$

Güvenirlik indeksi tek değişkenlik kaynaklı tesadüfi desende KTK'nın Cronbach Alpha katsayısına hemen hemen eşittir (Brennan, 2001).

Brennan (2004), G ve Φ katsayılarının yeterli ölçütlerinin isteğe bağlı olarak değiştiğini ancak bazı araştırmacıların G ve Φ katsayılarının 0,80'den büyük olması durumunda "yüksek" olarak değerlendirilebileceğini ifade etmektedir (Akt: Atılgan, 2005b). Bunun yanında Shavelson ve Webb (1991) ise

G ve Φ katsayısının en az 0,80 olması gerektiğini belirtmişlerdir.

G ve Φ katsayılarında yer alan hata varyanslarını azaltabilmek için alternatif K çalışmaları düzenlenebilir. Klasik Test Kuramında Sperman-Brown tahmin katsayısında ya da örneklem kuramında ortalamanın standart hatasında olduğu gibi nm ve nd ile gösterilen madde ve değerlendirici sayıları yerine değişik kombinasyonlarda değerlendirici ve madde sayıları yazılarak hata varyanslarının azaltılması ve daha yüksek G ve Φ katsayılarına ulaşmanın alternatifleri araştırılabilir. Bu tür bir araştırmaya G kuramında alternatif K çalışmaları adı verilir (Shavelson, 2003).

Çok değişkenli bir Genellenebilirlik deseninde alt testler sabit değişkenlik kaynağı olarak dikkate alınır. Her sabit değişkenlik kaynağı içindeki elemanların sayısı farklı alınır. Böyle bir durumda çok değişkenli G kuramı çalışması ile her bir sabit alt test için varyans bileşenlerinin ve dereceleme ölçeği puanların G ve Φ katsayılarının hesaplanabilmesi olanaklıdır (Brennan, 2001; Crocker ve Algina, 1986; Shavelson ve Webb, 1991).

G kuramının, KTK'na göre avantajları aşağıda sıralanmıştır.

- G kuramı bir ölçme durumunda yer alan bütün potansiyel hata kaynaklarını birlikte ve eşzamanlı olarak değerlendirerek, tek bir hata kaynağını değerlendiren KTK'na göre daha kapsamlı bir güvenilirlik kestirimi yapılmasına olanak sağlar.

- G kuramı ölçmenin güvenilirliğin belirlenmesinde mutlak hem de bağıl değerlendirmeler için KTK'da ise sadece bağıl değerlendirme için güvenilirlik hesaplanır.

- G kuramı geçerlik ve güvenilirlik arasındaki geleneksel farklılığı da bir ölçüde ortadan kaldırmaktadır. Genellenebilirlik kuramında kabul edilebilir gözlemler evreninden alınan örneklemin evrene genellenebilirliği test edildiğinden, kapsam geçerliğinin de bir kanıtı olarak kabul edilebilmektedir.

- G kuramı, çoklu hata kaynaklarını birleştirdiği ve test-tekrar test

güvenirliđi, iç tutarlılık ve puanlayıcılar arası güvenilirliklerin birlikte deđerlendirmektedir (Brennan, 2001; Yin ve Shavelson, 2004).

İlgili Araştırmalar

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

Yurt İçinde Yapılan Araştırmalar

Atılğan (2004) "Genellenebilirlik Kuramı ve Çok Değişken Kaynaklı Rasch Modelinin Karşılaştırılmasına İlişkin Bir Araştırma" adlı çalışmasında, gerçek verilerle 2002 ve 2003 yıllarında Müzik öğretmenliği özel yetenek seçme sınavı verilerini kullanarak Genellenebilirlik kuramı ve Çok Değişken Kaynaklı Rasch Modeli analizlerini yapmıştır. Analizler sonucunda elde edilen verilerle G kuramı ve Çok Değişken Kaynaklı Rasch Modeli istatistikleri karşılaştırılmıştır. Elde edilen sonuçlara göre, G kuramının tek değişkenli ve çok değişkenli modellerinin aynı ölçme durumu için alt boyutlardan oluşan testlerde farklı sonuçlar ürettiği görülmektedir. Ayrıca G ve Phi katsayılarının modelden etkilendiği ve alternatif karar çalışmalarıyla farklı puanlayıcı sayıları senaryoları karşısında elde edilen G ve Phi katsayılarının gerçek durumda kestirilenlerden farklı olduğu görülmüştür. Bununla birlikte hem G Kuramı hem de Çok Değişken Kaynaklı Rasch Modeli yaklaşımlarının her ikisinin de oldukça kullanışlı sonuçlar ürettiği görülmektedir.

Yelboğa (2007) "KTK ve G Kuramına Göre Güvenirliğin bir İş Performansı Ölçeği Üzerinde incelenmesi" adlı çalışmasında, 2005 ve 2006 yıllarında iş performansı ölçeği kullanılarak, KTK ve G Kuramı analizlerini gerçek verilerle kullanılmıştır. KTK'nda puanlama tekrar puanlama ve Cronbach alfa katsayıları, Kendall'ın Konkordans katsayısı ve G Kuramında ise çok değişkenli modelle G ve Phi katsayıları hesaplanmıştır. Hesaplanan bu katsayılar karşılaştırılmıştır. Elde edilen sonuçlara göre KTK ve G Kuramı güvenilirlik katsayıları birbirleriyle uyumlu sonuçlar ürettiği görülmüştür. Bununla birlikte güvenirlüğün iş performansı ölçeğinde değerlendirilmesinde G Kuramının kullanılmasının uygun olacağı

sonucuna varılmıştır. Aynı zamanda araştırmada kullanılan iş performansı ölçeğinin geçerli ve güvenilirliği yüksek bir ölçme aracı olarak değerlendirilebileceği ifade edilebilir.

Atılğan (2008)'in araştırmasında, 2003-2004 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 (deşifre; 4 görev, söyleme; 5 görev ve çalma; 4 görev) dört puanlayıcı tarafından puanlanmıştır. Sonuç 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 dereceleme ölçeği ve alt boyutların G ve Phi katsayıları hesaplanmıştır. Boyutlara göre G katsayıları; deşifre 0,87, söyleme 0,80 ve çalma 0,89 olarak bulunmuştur. Ölçeğin ise G katsayısı 0,90 olarak hesaplanmıştır. Boyutlara göre Phi katsayıları ise; deşifre 0,85, söyleme 0,51 ve çalma 0,82 olarak bulunmuştur. Dereceleme ölçeğin G katsayısı 0,76 olarak 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.

Genel olarak yurt içindeki araştırma bulguları değerlendirildiğinde, KTK ile G Kuramının güvenilirlik katsayıları birbirleriyle uyumlu sonuçlar ürettiği ve G Kuramının kullanılmasının daha uygun olacağı şeklinde değerlendirilmiştir. Alternatif K Çalışmaları ile G ve Phi katsayılarının artırılması için puanlayıcı sayısının artırılmasından çok madde sayılarını artırmanın daha uygun olacağı sonucuna varılmıştır.

Yurt Dışında Yapılan Araştırmalar

Tobar ve ark., (1999), 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 desene 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), 1992’de uygulanan Iowa Temel Yetiler Testi (Iowa Tests of Basic Skills) ve Iowa Eğitimsel Gelişme Testi (Iowa Tests of Educational Development) K formunun ulusal standartlaştırma çalışması verilerinden çekilen tesadüf bir örneklemin verileri kullanılmıştır. Araştırmada, test puanlarının güvenilirliği için hesaplanan Cronbach alfa güvenilirlik katsayısı ile G katsayısı arasındaki farkı belirlemek amaçlanmıştır. Araştırma sonucunda, kestirilen G katsayısı ve Cronbach alfa katsayısı arasındaki fark 0,04 olarak bulunmuştur. Ayrıca G katsayısının, test içindeki pasajların sayısının değişmesi ve toplam madde sayısının sabit tutulması durumundan, pasaj sayısı ile toplam madde sayısının sabit tutulması ve madde sayısının değiştirilmesi durumlarından daha çok değiştiği, bu nedenle her bir pasaj içindeki madde sayısının değiştirilmesi yerine pasaj sayısının değiştirilmesinin daha etkili olduğu sonucuna varılmıştır.

Johnson ve ark., (2000), dereceli puanlama anahtarı (rubric) ile yaptıkları deneysel bir çalışmada 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. Araştırmanın verileri olarak kullanılan 120 kompozisyon, Georgia Lisesi Yazma testinden alınmıştır.

Bir çift puanlayıcının ve bir çift 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.

Goodwin (2001), KTK yaklaşımını kullanarak puanlayıcılar arası uyum ve güvenilirliği araştırmıştır. Araştırmada, 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 (ortancası %20) ve puanlayıcıların uyumu % 60-100 aralığında (ortancası % 80) bulunurken, Kappa istatistikleri 0 ile 0,63 arasında hesaplanmıştır. G kuramı yaklaşımıyla hesaplanan G katsayısı ise 0,88 olarak bulunmuştur. Bu üç tekniğin hipotetik veriler üzerinde uygulanmasından elde edilen bulgulardan hareketle yetilerin avantaj ve dezavantajları tartışılmış, birçok hata kaynağını ayırabilmesi, daha kapsamlı ve esnek olması nedeniyle G kuramını araştırmalarda daha avantajlı olduğu vurgulanmıştır.

Lee ve ark., (2002), yeni TOEFL sınav sisteminin yazma ve konuşma bölümlerinin puanlarının güvenilirliğini Educational Testing Service (ETS)'de yaptıkları bir çalışmada G kuramı ile araştırmışlardır. Araştırmanın birinci aşamasında 488 kişiye yöneltilen iki okuma-yazma ve üç dinleme-yazma görevi 27 kişilik puanlayıcı grubundan iki puanlayıcı tarafından beş kategorili bir ölçek üzerinden puanlanmıştır. Ayrıca üç bağımsız yazma görevi, üç farklı alt gruba uygulanmış, böylece her bir grupta beş bağımsız yazma görevi verisi elde edilmiştir. İkinci aşamada üçüncü alt grubun, altı göreve ilişkin yazdıkları yazı her biri aynı altı puanlayıcı tarafından puanlanmıştır. G kuramında birey (b), görev (g) ve puanlayıcı (p) olmak üzere $b \times g \times p$ deseniyle incelenmiştir. Üçüncü aşamada konuşma yetilerine ilişkin olarak; altı dinleme-konuşma, iki okuma-

konuşma ve beş bağımsız konuşmadan oluşan görevler 261 adaya yöneltilmiş ve bireylerin tepkileri beş kategorili bir ölçekte 22 kişilik bir puanlayıcı havuzundan bağımsız iki puanlayıcı tarafından puanlanmıştır. Analizler sonucunda, her bir görev için hem yazma hem de konuşma bölümlerinde puanlayıcı sayısının artırılmasının ek maliyet getirmesinden dolayı görev sayısının artırılmasının puanların güvenilirliğini artırdığı sonucuna varılmıştır.

O'Brian ve ark., (2003), kekeme yetişkinlerde konuşma doğallığının ölçülmesinde kullanılan 9 kategorili "Konuşma Doğallığı Ölçeği" (Speech Naturalness Scale) ile G kuramını kullanarak bir araştırma yapmışlardır. Sidney üniversitesindeki akademik personel olmayan 15 puanlayıcı, konuşma tedavi programına devam eden 10 kekeme ve yaş-cinsiyet değişkenleri bakımından eşleştirilmiş 10 kişilik kontrol grubu konuşmacısını puanlamışlardır. Video kayıtlarını, grupları ve tedavi öncesi ya da sonrası olduğunu bilmeksizin 9 kategorili konuşma doğallığı ölçeğinin puanlanmasından elde edilen veriler araştırmada kullanılmıştır. Araştırma sonucunda, G kuramı tekniklerinin gözlem verilerinin güvenilirliğinin analizinde kullanımının ve güvenilirliğin ekonomik olarak sağlanabilmesi bakımından konuşma ve dil alanında yapılacak araştırmalarda kullanılabilir ve avantajlı bir yöntem olduğu belirtilmiştir (Akt: Atılğan, 2004).

Genel olarak yurt dışındaki araştırma bulguları değerlendirildiğinde, G Kuramı ve KTK güvenilirlik katsayılarının karşılaştırıldığı görülmüştür. Puanlayıcı sayısının artırılmasının puanların güvenilirliğini artırdığı sonucuna varılmıştır. Birçok hata kaynağını ayırabilmesi, daha kapsamlı ve esnek olması nedeniyle G kuramını araştırmalarda daha avantajlı olduğu vurgulanmıştır.

Genel olarak yapılan araştırmalara bakıldığında G kuramının kullanıldığı araştırmaların, ölçme araç ve yöntemlerinin özellikleri ve kuramsal özelliklerin incelenmesi ile ilgili olduğu görülmektedir.

BÖLÜM III

YÖNTEM

Bu başlık altında araştırmanın modeli, çalışma grubu, veri toplama araçları, verilerin toplanması ve verilerin analizi ile ilgili açıklamalar özetlenmiştir.

Araştırmanın Modeli

Araştırma, psikomotor yetilerin değerlendirilmesi çerçevesinde, puanlayıcıların tutarlılığının hesaplanmasında Klasik Test Kuramı ve Genellenebilirlik kuramının avantaj ve dezavantajlarının ortaya konmasına, kuramlardan hangisinin daha çok bilgi verdiği saptanmasına ve iki kurama dayalı olarak elde edilen güvenilirlik katsayılarının karşılaştırmasına dayalıdır. Bu yönüyle karşılaştırmalı, açıklama düzeyinde kuramsal bir araştırmadır.

Çalışma Grubu

Bu araştırmanın çalışma grubunu, 2007-2008 yılları arasında Ankara ilindeki profesyonel spor kulüplerinin futbol takımlarının alt yapılarındaki profesyonelliğe aday futbolcu statüsünde bulunan toplam 72 futbolcu (ortalama-yaş= 17,13±1,15 yıl; antrenman-yaş= 6,26±1,12 yıl) oluşturmuştur. Futbol branşlarındaki üst düzey profesyonel futbol takımlarından belirlenmiş bir grup ile çalışılmıştır. Alt yapıları oluşturan futbolcuların ailelerinden, kendilerinden ve takımların antrenörlerinden yetilerin ölçülmesi için önceden oluşturulmuş formlarla izinler alınmıştır. Futbolcuların yeti düzeyleri arasındaki farklılığı elimine etmek amacıyla çalışma grubunu antrenman yılları 5 ve üzeri olan futbolcular oluşturmuştur.

Veri Toplama Araçları

Bu araştırmada; futbolcuların yetilerinin (pas, şut çekme, top sürme ve top kontrolü) ölçülmesi amacıyla Luxbacher (1991)'in 56 maddeden oluşan "Futbol Yetilerine İlişkin Dereceleme Ölçeği" kullanılmıştır (Ek 1).

Futbol Yetilerine İlişkin Dereceleme Ölçeğine ait maddeler için, futbolda bazı yetileri kapsayan ve ölçeğin alt boyutlarını oluşturan pas (k=17), top kontrolü (k=11), top sürme (k=12) ve şut, (k=16) yetilerine ait Worthington (1974) ve Luxbacher (1991) tarafından belirtilen kriterler (maddeler) incelenmiştir. Futbol yetilerine ait maddeler kapsam açısından incelendiğinde Luxbacher (1991)' e ait maddeler uygun bulunmuş ve dereceleme ölçeğinde kullanılmıştır. Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin alt boyutlarına ait maddeler bir İngilizce okutmanı ve beden eğitimi ve spor bölümünde öğretim üyesi olan iki uzman tarafından Türkçe'ye çevrilmiştir. Elde edilen iki Türkçe ölçekteki maddeler karşılaştırılmış ve aynı çeviriye sahip maddeler belirlenmiştir. Çeviriler aynı uzmanlar tarafından tekrar İngilizceye çevrilmiştir. İngilizce çeviri orijinal ölçek ile karşılaştırılarak birbirine en yakın çeviriler ile dereceleme ölçeğin Türkçe şekli oluşturulmuştur.

Türkçe'ye çevrilmiş "Futbol Yetilerine İlişkin Dereceleme Ölçeği" futbol branşındaki iki akademisyen ve ölçme ve değerlendirme konusunda bir uzman tarafından incelenmiş ve ölçek maddeleri önerilere göre düzeltilerek nihai form haline getirilmiştir. Bu inceleme sırasında ölçek madde sayıları değişmemiştir. Böylece dereceleme ölçeğinin, ölçülmek istenilen futbol yetilerini kapsayıp kapsamadığı uzman görüşlerine dayalı olarak belirlenmiştir.

Performansa ait yetilerin hiyerarşik bir sırayla yapılmasının gerekliliğinden dolayı Futbol Yetilerine İlişkin Dereceleme Ölçeği'ne ait maddelerin sıraları değişmemiştir. Dereceleme ölçeğinden alınabilecek en düşük puan 0 ve en yüksek puan 224 arasında değişmektedir. Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin alt boyutlarından alınabilecek en düşük-en yüksek puanlar incelendiğinde pas yetisi 0-68, Top kontrol yetisi 0-44, Top sürme yetisi 0-48 ve şut çekme yetisi 0-64 arasında değişmektedir. Futbol Yetilerine İlişkin Dereceleme Ölçeğine ait maddelerde belirtilen yetilerden, beş kategorili ölçekte puanlayıcılardan "Görülmedi", "Yetersiz", "Yeterli", "Oldukça Yeterli" ve

“Mükemmel” seçeneklerinden birini işaretlemeleri istenmiştir. Ölçülmek istenilen yetiyi tam anlamıyla gerçekleştiren futbolcunun alacağı “Mükemmel” derecelemesi “4” ve futbolcu tarafından yetinin gerçekleştirilmemesi durumunda alacağı “Görülmedi” ibaresine karşılık gelen “0” arasında değişen puanlar belirlenmiştir. Ölçeğin bütünü ve alt boyutlarına ait maddelerden alınan puanların yükselmesi, futbol yetisinin yüksek olduğunu ifade etmektedir.

Verilerin Toplanması

Araştırmanın amacına uygun olarak yetilere ait ölçmeler çalışma grubuna katılan farklı yaşlardaki futbolcuların sezon içerisindeki müsabaka dönemindeki antrenmanlarından önce futbolcuların antrenman kıyafetleriyle takımdaki kadro sırasına uygun olarak alınmıştır. Futbolcular, çim futbol sahasında sırasıyla ayak içi pas, ayak içi top kontrolü, top sürme ve ayak üstü şut yetilerini 5'er kez uygulamıştır. Yetilerin uygulanması sırasında video-kamera ile elde edilen görüntüler, yetilerdeki bütün aksiyonların rahat görünebileceği şekilde uygulayıcıların baskın olan ayağının olduğu taraftan daha önceden video kamera çekimi konusunda eğitimi verilmiş uzman kişi tarafından 3 metre uzaklıktan çekilmiştir.

Dereceleme ölçeğindeki yetiler puanlayıcılar tarafından incelenmiş ve maddelerin puanlanması sırasında dikkat edilmesi gereken kriterler belirlenmiştir. Elde edilen görüntüler antrenörlük yılları 2 ve üzeri olan B lisans futbol antrenörlük belgesine sahip 4 uzman tarafından değerlendirilmiştir. Puanlayıcıların ayrı ortamlarda, dereceleme ölçeğindeki futbol yetilerine ait maddeleri tam anlamıyla görebilmelerine imkân sağlanması amacıyla % 50 oranında yavaşlatılmış görüntüleri bilgisayardan izlemeleri ve puanlamaları istenmiştir. Puanlayıcılar görüntüleri aynı gün içerisinde, birer saatlik oturumlarda ve oturumlar arasında 30 dakika dinlenme olacak şekilde izleyerek puanlamışlardır. Puanlayıcılar, birbirlerinin bilgisayar ekranlarını ve puanlama yaptıkları dereceleme ölçeklerini göremeyecekleri şekilde sınıf ortamında puanlamalarını yapmışlardır.

Puanlayıcıların cevaplarındaki tutarlılıklarını tespit etmek amacıyla Futbol Yetilerine İlişkin Dereceleme Ölçeğine ait maddelerde belirtilen yetiler bakımından aynı futbolcular bir hafta sonra tekrar ilk puanlamadaki uygulamaya benzer olarak puanlanmıştır.

Verilerin Analizi

Puanlayıcılar, araştırma kapsamında profesyonel futbol takımlarının alt yapısını oluşturan futbolcuları Futbol Yetilerine İlişkin Dereceleme Ölçeği kapsamındaki futbol yetileri bakımından bir hafta arayla iki kez puanlamışlardır. Bu puanlamalardan ölçülen özelliğe daha aşina olmaları bakımından ikinci puanlamaya ait veriler analizlerde kullanılmıştır.

Tekrarlanan puanlama işleminde hata paylarının daha az olabileceği kabul edilerek araştırmada, puanlayıcıların Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin ikinci puanlamasına ait maddelerden elde edilen veriler çözümlenmiştir.

Araştırmanın amaçları doğrultusunda Futbol Yetilerine İlişkin Dereceleme Ölçeği'nden elde edilen veriler ile ilk puanlama ve ikinci puanlama arasındaki tutarlılık için "Pearson momentler çarpımı" korelasyon katsayısı bulunmuştur. Futbol yetilerine ait maddelerin kendi içinde tutarlı ölçme yapıp yapmadığının belirlenmesi amacıyla maddelerin iç tutarlılık güvenirliği için Cronbach α (alfa) katsayısı hesaplanmıştır. Puanlayıcıların Futbol Yetilerine İlişkin Dereceleme Ölçeğini puanlanmaları arasındaki tutarlılık için Kendall uyum katsayıları hesaplanmıştır. Pearson momentler çarpımı, Cronbach α ve Kendall'in uyum katsayıları için SPSS 10.0 programı kullanılmıştır.

Futbol Yetilerine İlişkin Dereceleme Ölçeği ve alt boyutlarından elde edilen puanların güvenirliği, puanlayıcı sayılarının 3, 4 ve 5 olduğu ve madde sayılarının bir arttırılıp bir azaltıldığı koşullarda, G kuramında temel ve ortak etkilerin varyans bileşenlerinin kestirilmesi için MGENOVA paket programında 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. Madde (m) (görev) temel etkisi için kestirilen varyans bileşenin yüksek olması bazı maddelerin diğerlerine göre madde güçlük düzeylerinin daha yüksek olduğunu, sıfıra yakın olması madde güçlük düzeylerinin birbirine yakın olduğunu göstermektedir. Puanlayıcı (p) temel etkisi için varyans bileşenin yüksek olması bazı puanlayıcıların diğerlerine göre daha katı ya da cömertlikte puanlama yapması, sıfıra yakın olması ise puanlayıcıların birbirine benzer puanlama yaptığını göstermektedir (Shavelson ve Webb,1991).

Birey-madde ortak etkisi ($b*m$); bireyin bağıl konumunun bir maddeden diğerine değişip değişmediğini bir başka deyişle, belli bir maddenin, bütün 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 maddeden diğer maddeye bağıl konumlarının değiştiği şekilde yorumlanabilir. Birey-puanlayıcı ortak etkisi ($b*p$) belli bir puanlayıcının, bütün bireyler için aynı biçimde puanlama yapıp yapmadığı ya da belli bir bireyin puanlanmasıyla diğer bireylerin puanlaması 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. Madde-puanlayıcı ortak etkisi ($m*p$) belli bir puanlayıcının bütün maddeler için aynı biçimde ya da belli bir görev için diğerlerinden farklı davranıp davranmadığını, puanlayıcıların puanlamalarının bir maddeden diğerine kararlı olup olmadığını ya da etkilenip etkilenmediğini göstermektedir. Kestirilen varyans bileşenin sıfır ya da sıfıra yakın olması durumunda puanlayıcıların bir maddeden diğerine puanlamalarında kararlı olduklarını göstermektedir (Shavelson ve Webb,1991).

Kalan etki varyansı (b^*m^*p,e) Birey-madde-puanlayıcı ortak etkisi ve ölçülemeyen varyans kaynağı, tesadüfî hata ya da sabit hatayı göstermektedir. Varyans bileşen değerinin sıfıra yakın olması bxm^*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).

Verilerin analizi sonucunda elde edilen bulgular, araştırma soruları çerçevesinde yorumlanarak tartışılmıştır.

BÖLÜM IV

BULGULAR VE YORUMLAR

Bu bölümde, Futbol Yetilerine İlişkin Dereceleme Ölçeği ve alt boyutlarından elde edilen puanlamalara göre araştırmanın problemlerinden elde edilen bulgular ve bulgulara dayalı olarak yapılan yorumlara yer verilmiştir.

Bulgular ve yorumlar her bir alt amaca göre açıklanmadan önce araştırmada kullanılan Futbol Yetilerine İlişkin Dereceleme Ölçeği ile alt boyutlarına ilişkin toplam puan ve kategori olarak bakıldığında toplam puanın madde sayısına bölünmesiyle elde edilen ortalama madde puanlarının ilk ve ikinci puanlamalarından elde edilen betimsel istatistikler Tablo 3’de verilmiştir.

Tablo 3. Futbol Yetilerine İlişkin Dereceleme Ölçeği’nin İlk ve İkinci Puanlamaya Ait Betimsel İstatistikler

İlk Puanlama	Puan Aralığı	Ortalama	SS	Mod	Medyan	Çarpıklık	Basıklık
Genel Puanlama	56-224 (0-4)	142,26 (2,54)	14,22 (0,19)	2,35	2,57	-0,30	-1,078
Pas Yetisi	17-68 (0-4)	48,28 (2,84)	6,96 (0,31)	2,32	2,94	-0,559	-0,921
Top Kontrol Yetisi	11-44 (0-4)	27,84 (2,53)	4,33 (0,27)	2,68	2,53	-0,201	-0,439
Top Sürme Yetisi	12-48 (0-4)	25,44 (2,11)	4,33 (0,19)	1,90	2,12	-0,028	-0,759
Şut Yetisi	16-64 (0-4)	40,70 (2,54)	6,06 (0,30)	2,56	2,53	0,262	-0,692
İkinci Puanlama	Puan Aralığı	Ortalama	SS	Mod	Medyan	Çarpıklık	Basıklık
Genel Puanlama	56-224 (0-4)	147,27 (2,62)	11,24 (0,15)	2,64	2,65	-0,233	-0,962
Pas Yetisi	17-68 (0-4)	49,09 (2,88)	5,70 (0,27)	3,01	2,95	-0,708	-0,473
Top Kontrol Yetisi	11-44 (0-4)	28,67 (2,60)	2,97 (0,19)	2,75	2,63	-0,103	-0,338
Top Sürme Yetisi	12-48 (0-4)	26,71 (2,22)	3,93 (0,17)	2,27	2,25	-0,290	-0,141
Şut Yetisi	16-64 (0-4)	42,81 (2,67)	6,00 (0,24)	2,44	2,63	0,441	-0,780

Tablo 3 incelendiğinde Futbol Yetilerine İlişkin Dereceleme Ölçeği ile alt boyutlarına ilişkin ilk puanlamanın ortalama verileri, ikinci puanlama ortalama verilerinden daha düşük olduğu görülmektedir. Her iki puanlama incelendiğinde çarpıklık katsayısının +1 ile -1 arasında sınırlı kalması dağılımın normale yakın bir dağılım gösterdiği şeklinde yorumlanabilir. Bundan dolayı tüm puan türlerinde grubun normal dağılıma yakın bir dağılım gösterdiği, şut yetisi dışındaki puanlamaların ise sola çarpık olduğu görülmektedir. Normal dağılıma göre basıklık katsayıları incelendiğinde şut yetisi dışında ikinci puanlama verilerinin, ilk

puanlama verilerinden daha basık bir dağılım gösterdiği saptanmıştır (Şencan, 2005; Turgut, 1997).

1) Klasik Test Kuramına göre;

a. Futbol Yetilerine İlişkin Dereceleme Ölçeğine ilişkin puanlamalar arası tutarlılığın derecesi nedir?

Futbol Yetilerine İlişkin Dereceleme Ölçeği'nden elde edilen veriler ile ilk puanlama ve ikinci puanlama arasındaki tutarlılık için Pearson momentler çarpımı korelasyon sonuçları bu bölümde tartışılmıştır.

Futbol Yetilerine İlişkin Dereceleme Ölçeği ile alt boyutlara ilişkin puanlayıcıların, aynı kişileri ilk ve ikinci puanlamaları ile elde edilen sonuçlar arasındaki tutarlılık düzeyi Tablo 4'de verilmiştir.

Tablo 4. Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin İlk ve İkinci Puanlamalar Arası Güvenirlik Katsayıları

Boyutlar	İlk ve ikinci Puanlama Güvenirliği
Pas Yetisi (k=17)	0,87
Top Kontrol Yetisi (k=11)	0,82
Top Sürme Yetisi (k=12)	0,73
Şut Yetisi (k=16)	0,75
Ölçek Toplam Puan (k= 56)	0,80

Tablo 4 incelendiğinde Futbol Yetilerine İlişkin Dereceleme Ölçeği içerisindeki 56 madde için ilk puanlama ve ikinci puanlama uygulaması sonucunda Pearson momentler çarpımı ile hesaplanan korelasyon katsayısı 0,80 olarak bulunmuştur. Alt boyutlar için yapılan Pearson momentler çarpımı sonuçlarına göre; güvenilirlik katsayıları 0,73 ile 0,87 arasında değişmektedir. İlk ve ikinci puanlama sonucunda bulunan değere göre ölçeğin yüksek kararlılık

gösterdiği görülmektedir (Szymanski ve Linkowski, 1993).

Puanlayıcıların ölçeği ve alt boyutları puanlamalarından elde edilen yüksek güvenirlik katsayıları, yetilerin bütün puanlayıcılar tarafından kolayca görüldüğü şeklinde yorumlanabilir.

Dört puanlayıcının, aynı kişileri ilk ve ikinci puanlamaları ile elde edilen sonuçlar arasındaki tutarlılık düzeyi Tablo 5’de verilmiştir.

Tablo 5. Futbol Yetilerine İlişkin Dereceleme Ölçeğine İlişkin Dört Puanlayıcının İlk ve İkinci Puanlamalar Arası Güvenirlik (Tutarlılık) Katsayıları

Puanlayıcı I	Boyutlar	Puanlamalar arası Güvenirlik	Ölçek Toplam Puan ait puanlamalar arası güvenirlik
1. Puanlayıcı	Pas Yetisi	0,88	0,81
	Top Kontrol Yetisi	0,73	
	Top Sürme Yetisi	0,67	

	Şut Yetisi	0,87	
2. Puanlayıcı I	Pas Yetisi	0,83	0,82
	Top Kontrol Yetisi	0,75	
	Top Sürme Yetisi	0,67	
	Şut Yetisi	0,84	
3. Puanlayıcı I	Pas Yetisi	0,87	0,83
	Top Kontrol Yetisi	0,71	
	Top Sürme Yetisi	0,75	
	Şut Yetisi	0,72	
4. Puanlayıcı I	Pas Yetisi	0,77	0,77
	Top Kontrol Yetisi	0,68	
	Top Sürme Yetisi	0,69	
	Şut Yetisi	0,80	

Tablo 5 incelendiğinde puanlayıcıların Futbol Yetilerine İlişkin Dereceleme Ölçeği içerisindeki 56 madde için ilk puanlama ve ikinci puanlama uygulaması sonucunda Pearson momentler çarpımı ile hesaplanan korelasyon katsayısı 0,77 ile 0,83 arasında değiştiği görülmektedir. Alt boyutlar için dört puanlayıcının puanlamalarına ilişkin yapılan Pearson momentler çarpımı güvenilirlik katsayıları 0,67 ile 0,88 arasında değiştiği görülmektedir.

Puanlayıcıların alt boyutları puanlamalarında bazı boyutlarda düşük güvenilirlik katsayıları elde edilmiştir. Alt boyutlara ilişkin düşük güvenilirlik katsayıları yetinin puanlayıcı tarafından zor görüldüğü şeklinde yorumlanabilir. Puanlayıcıların dereceleme ölçeğinin bütününe ait puanlamaları incelendiğinde ilk ve ikinci puanlama sonucunda bulunan değerlere göre ölçeğin yüksek kararlılık gösterdiği görülmektedir (Szymanski ve Linkowski, 1993).

b. Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin toplam ve alt boyutlarına ait iç tutarlılığın derecesi nedir?

Futbol Yetilerine İlişkin Dereceleme Ölçeği'nden elde edilen puanlamalara göre maddelerin iç tutarlılık güvenilirliği olarak kestirilen Cronbach α (alfa) değerleri Tablo 6'da verilmiştir.

Tablo 6. Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin Cronbach Alfa İç tutarlılık Güvenirliği

Boyutlar	Cronbach Alfa
Pas Yetisi	0,93
Top Kontrol Yetisi	0,81
Top Sürme Yetisi	0,71
Şut Yetisi	0,89
Ölçek Toplam Puan	0,90

Tablo 6 incelendiğinde Futbol Yetilerine İlişkin Dereceleme Ölçeği içerisindeki 56 madde için puanlamalara ait güvenilirlik katsayısı Cronbach α =0,90 olarak bulunmuştur. Alt boyutlara ait güvenilirlik katsayıları incelendiğinde 0,71 ile 0,93 arasında değişen Cronbach α katsayıları elde edilmiştir.

Futbol Yetilerine İlişkin Dereceleme Ölçeğini oluşturan maddelerin puanlamalarına ilişkin iç tutarlılık katsayılarının saptanması için hesaplanan Cronbach α katsayısı için en az 0,70 (Nunnally, 1978) genel kabulüne göre maddeler büyük ölçüde birbirleriyle tutarlı ve aynı özelliğin öğelerini yoklayan maddelerden oluştuğu şeklinde yorumlanabilir (Akt: Tavşancıl, 2002).

Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin puanlayıcıların toplam ölçek ve alt ölçekteki Cronbach alfa (α) iç tutarlılık güvenilirlik değerleri Tablo 7'de verilmiştir.

Tablo 7. Puanlayıcıların Puanlamalarına Ait Toplam Ölçek ve Alt Ölçekteki Cronbach Alfa İç Tutarlılık Güvenirliği

Puanlayıcı	Boyutlar	Cronbach Alfa	Ölçek Toplam Puan ait Cronbach Alfa
1. Puanlayıcı	Pas Yetisi	0,95	0,93
	Top Kontrol Yetisi	0,87	
	Top Sürme Yetisi	0,81	
	Şut Yetisi	0,92	
2. Puanlayıcı	Pas Yetisi	0,93	0,91
	Top Kontrol Yetisi	0,81	

	Top Sürme Yetisi	0,81	
	Şut Yetisi	0,90	
3. Puanlayıcı	Pas Yetisi	0,90	0,77
	Top Kontrol Yetisi	0,57	
	Top Sürme Yetisi	0,69	
	Şut Yetisi	0,84	
	Pas Yetisi	0,84	
4. Puanlayıcı	Top Kontrol Yetisi	0,65	0,82
	Top Sürme Yetisi	0,60	
	Şut Yetisi	0,81	

Tablo 7 incelendiğinde Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin her bir puanlayıcının puanlamalarına ilişkin Cronbach α katsayılarına bakıldığında; Cronbach α güvenilirlik katsayıları 0,77 ile 0,93 arasında değiştiği görülmektedir.

Puanlayıcıların alt boyutları puanlamaları sonucunda bazı boyutlarda düşük iç tutarlılık katsayıları elde edilmiştir. Alt boyutlara ilişkin iç tutarlılık katsayılarının düşük olması bazı yetilere ait maddelerin puanlayıcılar tarafından tutarlı puanlanmadığı şeklinde yorumlanmaktadır. Futbol Yetilerine İlişkin Dereceleme Ölçeğini oluşturan maddelerin her bir puanlayıcının puanlarına ait Cronbach α iç tutarlılık katsayıları en az 0,70 (Nunnally, 1978) genel kabulüne göre incelenmiştir (Akt: Tavşancıl, 2002). Böylece ölçeğin bütününe ait maddeler büyük ölçüde birbirleriyle tutarlı ve aynı özelliğin öğelerini yoklayan maddelerden oluştuğu şeklinde yorumlanabilir.

Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin puanlanmasında, puanlayıcılar arasındaki tutarlılığın test edilmesi için hesaplanan Kendall uyum katsayıları (W) bu bölümde tartışılmıştır.

c. Futbol Yetilerine İlişkin Dereceleme Ölçeğine göre yapılan puanlamada puanlayıcılar arası tutarlılığın derecesi nedir?

Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin dört farklı puanlayıcının aynı koşullar altında bir kişiyi puanlamasıyla elde edilen puanlamalara göre elde

edilen puanlayıcılar arasındaki tutarlılık derecesi, parametrik olmayan istatistiksel bir teknik olan Kendall'ın uyuşum (konkordans) katsayısı ile analiz edilmiş ve bulunan değerler Tablo 8'de verilmiştir.

Tablo 8. Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin Dört Farklı Puanlayıcının Puanlamaları Arasındaki Kendall W Güvenirlik Katsayıları

Boyutlar	Boyutlar için Kendall W
Pas Yetisi	0,77
Top Kontrol Yetisi	0,72
Top Sürme Yetisi	0,80
Şut Yetisi	0,83
Ölçek Toplam Puan	0,78

Tablo 8 incelendiğinde Futbol Yetilerine İlişkin Dereceleme Ölçeği içerisindeki 56 madde için puanlayıcılar arası uyum katsayısı Kendall W =0,78 olarak bulunmuştur. Alt boyutlar için puanlayıcılar arası uyuma ait Kendall W katsayısı 0,72 ile 0,83 arasında değişim göstermektedir. Futbol Yetilerine İlişkin Dereceleme Ölçeğini oluşturan maddelere ait dört farklı puanlayıcının aynı koşullar altında bir kişiyi puanlamaları sonucunda dört farklı puanlayıcı arasındaki uyumun, Top sürme ve Şut yetisi dışındaki yetilerde 0,80 genel kabulüne göre düşük olduğu belirlenmiştir (Szymanski ve Linkowski, 1993).

2. Genellenebilirlik Kuramına göre;

a. Futbol Yetilerine İlişkin Dereceleme Ölçeği'nden elde edilen puanların, puanlayıcı sayılarının 3, 4 ve 5 olduğu ve madde sayılarının bir arttırılıp, bir azaltıldığı (4 boyut için ayrı olacak şekilde; k=60; k=52) koşullarda, kestirilen G ve Phi katsayılarının derecesi nedir?

Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin puanlamasından elde edilen veriler için G kuramının çok değişkenli modeline ait K çalışması sonucunda orijinal ölçek için hesaplanan G ve Phi katsayılarının, senaryolarla kestirilen G ve Phi katsayılarıyla tutarlı olup olmadığı bu bölümde incelenmiştir.

Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin puanlamasına ait çok değişkenli K çalışması sonucunda orijinal puanlayıcı-madde sayıları ile farklı puanlayıcı-madde sayıları senaryolarına göre kestirilen G ve Phi katsayıları Tablo 9'da verilmiştir.

Tablo 9. Futbol Yetilerine İlişkin Dereceleme Ölçeği Puanlamasının Çok Değişkenli K Çalışması Sonucunda Kestirilen G ve Phi katsayıları

Model	Ölçek Madde Sayıları	np=3		np=4		np=5	
		G	Φ	G	Φ	G	Φ
Çok Değişkenli	52	0,77887	0,72405	0,82329	0,77104	0,85246	0,80228
	56	0,78017	0,72785	0,82445	0,77468	0,85351	0,80578
	60	0,78133	0,73129	0,82548	0,77796	0,85446	0,80894

$$n^b = 72$$

Tablo 9 incelendiğinde her bir alt boyut için madde sayılarının birer artırılması ($n_m=60$) ve azaltılması ($n_m=52$) ile puanlayıcıların birer artırılması ($n_p=5$) ve azaltılması ($n_p=3$) senaryolarına göre G katsayılarının 0,77887 ile 0,85446 arasında ve Phi katsayılarının da 0,72405 ile 0,80894 arasında değiştiği görülmektedir.

Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin puanlama verilerine ilişkin genel olarak, dereceleme ölçeği için farklı puanlayıcı-madde senaryolarına göre kestirilen Phi katsayılarının G katsayılarından küçük olduğu görülmektedir. Bunun yanında G katsayılarının, farklı puanlayıcı-madde senaryolarına göre artışına ve azalışına paralel olarak, Phi katsayıları da paralel olarak artmakta ya da azalmakta olduğu görülmektedir.

Futbol Yetilerine İlişkin Dereceleme Ölçeğine ait çok değişkenli K çalışmasına göre kestirilen G ve Phi katsayılarından elde edilen bu bulgu, Johnson ve ark., (2000)'ın ve Atılgan (2004)'in çalışmalarındaki farklı puanlayıcı-madde senaryolarından elde edilen G ve Phi katsayılarının paralel olarak arttığı şeklindeki bulgularıyla tutarlılık göstermektedir.

G ve Phi katsayılarının elde edilen puanlama verilerine göre K çalışması sonucunda orijinal puanlayıcı-madde sayıları ile farklı puanlayıcı-madde sayıları senaryoları kestirilmiştir. Beş puanlayıcı 60 madde için en yüksek G ve Phi katsayılarına sahip olduğu görülmektedir. Bu bağlamda, ölçeğin orijinal hali ile farklı senaryolar karşısında kestirilen G ve Phi katsayılarının birbirlerine benzer sonuçlar verdiği dolayısıyla ölçeğin orijinal haliyle kullanılmasının uygun olabileceği sonucuna varılmıştır.

b. Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin puanlanmasından elde edilen verilerinin çok değişkenli G çalışması sonucunda alt boyutların kestirilen varyansları ve toplam varyansı açıklama oranları nedir?

Futbol Yetilerine İlişkin Dereceleme Ölçeği alt boyutlarının puanlanmasına ait varyans ve varyans yüzdelerini belirlemek amacıyla, G kuramının çok değişkenli bmxpx modeli kullanılarak varyans kaynakları ve toplam varyansı açıklama oranları elde edilmiştir. Çok değişkenli G çalışması için, kestirilen varyans bileşenleri ile toplam varyansı açıklama yüzdeleri birey (b), madde (m) ve puanlayıcı (p) temel etkileri, birey x madde (bxm), birey x puanlayıcı (bpx) ve madde x puanlayıcı (mxx) ortak etkileri ile artık varyans (bmxpx) ortak etkilerine ait sonuçlar bu bölümde verilmiştir.

Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin puanlanmasından elde edilen verilerin çok değişkenli G çalışması sonucunda alt boyutların kestirilen varyansları ve toplam varyansı açıklama oranları Tablo 10'da verilmiş ve alt boyutlar için çıkan sonuçlar sırasıyla yorumlanmıştır.

Tablo 10. Futbol Yetilerine İlişkin Dereceleme Ölçeğine İlişkin Puanlamada
bxmnp Modeli için Alt Boyutlarına ait Kestirilen Varyans ve Toplam Varyansı
Açıklama Oranları

Varyans Kaynağı	Çok değişkenli G Çalışması							
	Pas Yetisi		Top Kontrol Yetisi		Top Sürme Yetisi		Şut Çekme Yetisi	
	Varyans	%	Varyans	%	Varyans	%	Varyans	%
b	0,64820	75,566	0,71380	68,490	0,41830	52,338	0,63693	57,092
m	0,01513	1,764	0,04484	4,302	0,02125	2,659	0,08106	7,266
p	0,00486	0,567	0,01842	1,767	0,00533	0,667	0,04036	3,618
bm	0,00152	0,177	0,00958	0,919	0,00883	1,105	0,01633	1,464
bp	0,03043	3,547	0,02396	2,299	0,03846	4,812	0,06323	5,668
mp	0,01589	1,852	0,02195	2,106	0,01486	1,859	0,05752	5,156
bmp	0,14176	16,526	0,20965	20,116	0,29220	36,560	0,22020	19,738
Toplam	0,85779	100	1,0422	100	0,79923	100	1,11563	100

b: birey m: madde p: puanlayıcı

Tablo 10 incelendiğinde, Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin puanlanmasından elde edilen verilerin Pas yetisi alt boyutunda birey (b) temel etkisi için kestirilen varyans bileşenin en büyük varyans bileşeni (0,64820) olduğu ve toplam varyansın % 75,57'sini açıkladığı görülmektedir. Birey temel etkisi, bireyin ölçülen özellikleri bakımından farklılıklarını göstermesi bağlamında bireyler için kestirilen varyans bileşenin toplam varyans içinde en büyük payı alması bireylerin ölçülen performansları bakımından farklılıklarını ortaya koyabildiğinin bir göstergesidir.

Madde (m) temel etkisi için kestirilen varyans bileşeni (0,01513) toplam varyansın % 1,76'sını açıklamaktadır. Maddeler için kestirilen varyans bileşenin ve toplam varyansı açıklama oranının sıfıra yakın olması maddelerin güçlüklerinin

birbirlerine yakın olduğunu, bir maddenin diğerine göre daha zor olmadığını bir göstergesidir.

Puanlayıcı (p) temel etkisi için kestirilen varyans bileşeni (0,00486) toplam varyansın % 0,57'sini açıklamaktadır. Puanlayıcılar temel etkisi için kestirilen varyans bileşenin toplam varyansı açıklama yüzdesinin sıfıra yakın olması; puanlayıcıların bütün bireyler için yaptıkları puanlamaları arasında farklılık bulunmadığı şeklinde yorumlanabilir. Shavelson ve Webb (1991)'e göre puanlayıcı temel etkisi için kestirilen varyans bileşeninin, belli bir puanlayıcının diğerlerine göre bütün bireyler için verdikleri puanlarda daha cömert ya da katı davrandıklarının göstergesi olduğu şeklinde vurgulanmaktadır.

Birey x madde (bxm) ortak etkisi (0,00152) toplam varyansın % 0,18'ini açıklamaktadır. Birey x madde ortak etkisinin küçük olduğu ve belli bir bireyin bir maddeden diğer maddeye bağlı durumunun farklılık göstermediği şeklinde yorumlanmaktadır.

Birey x puanlayıcı (bpx) ortak etkisi (0,03043) toplam varyansın % 3,54'ünü açıklamaktadır. Birey x puanlayıcı ortak etkisi varyansının sıfıra yakın olması, belli bir puanlayıcının belli bir bireyi diğer puanlayıcılardan farklı puanlamadığı şeklinde yorumlanmaktadır.

Madde x puanlayıcı (mxx) ortak etkisi (0,01589) toplam varyansın % 1,85'ini açıklamaktadır. Madde x puanlayıcı ortak etki varyansı; puanlayıcıların bireyleri bir maddeden diğerine kararlı puanlayıp puanlamadıklarını gösterir. Madde x puanlayıcı ortak etkisi ve toplam varyansı açıklama oranının küçük olması, madde x puanlayıcı ortak etkisinden gelen farklarında küçük olduğunu, puanlayıcıların bireyleri belli bir maddeden diğerine aynı kararlılıkta puanladıkları biçiminde yorumlanabilir.

Pas yetisi alt boyutunda artık varyansın birey x madde x puanlayıcı ortak etkisi varyans bileşeninin (0,01416) toplam varyansı açıklama oranını % 16,52 olduğu görülmektedir. Artık varyansın büyük olması; bmxpx ortak etkisinden gelen farkların ve ölçmedeki tesadüfî hataların çokluğunun bir göstergesidir.

Futbol Yetilerine İlişkin Dereceleme Ölçeği puanlanmasından elde edilen verilere göre Top kontrolü yetisi alt boyutunda çok değişkenli analiz ile kestirilen en büyük varyans bileşeni, birey (b) temel etkisi için kestirilen varyans bileşeni (0,71380) toplam varyansın % 68,49'unu açıklamaktadır. Bireyler için kestirilen varyans bileşeninin ve toplam varyans içinde payının büyük olması, bireylerin ölçülen performansları bakımından aralarındaki farklılıkların ortaya konulabildiğinin bir göstergesidir.

Madde (m) temel etkisi için kestirilen varyans bileşeni (0,04484) toplam varyansın % 4,30'unu açıklamaktadır. Maddeler için kestirilen varyans bileşeninin ve toplam varyansı açıklama oranının sıfıra yakın olması maddelerin güçlüklerinin birbirlerine yakın olduğunun, bir maddenin diğerine göre daha zor olmadığına bir göstergesidir.

Puanlayıcı (p) temel etkisi için kestirilen varyans bileşeni (0,01842) ve toplam varyansın % 1,77'sini açıklamaktadır. G katsayısının hesaplanmasından kestirilen varyans değerinin sıfıra yakın olması; puanlayıcıların bütün bireyler için yaptıkları puanlamaları arasında farklılık bulunmadığı, puanlayıcıların bütün bireyleri birbirine yakın katılık/çömerlikte puanlar verdikleri söylenebilir.

Birey x madde (bxm) ortak etkisi (0,00958) toplam varyansın % 0,92'sini açıklamaktadır. Birey x madde ortak etkisinin bu boyutta, varyansı açıklama oranının küçük olduğu görülmektedir. Bu durum birey x madde ortak etkisinden kaynaklanan farklılıkların küçük olduğunu ve bireylerin bir maddeden diğerine durumlarının farklılaşmadığı şeklinde yorumlanmaktadır.

Birey x puanlayıcı (b_{xp}) ortak etkisi (0,02396) toplam varyansın % 2,30'unu açıklamaktadır. Birey x puanlayıcı ortak etkisi varyansının sıfıra yakın olması, belli bir puanlayıcının belli bir bireyi diğer puanlayıcılardan farklı puanlamadıkları şeklinde yorumlanmaktadır.

Madde x puanlayıcı (m_{xp}) ortak etkisi (0,02195) toplam varyansın % 2,11'ini açıklamaktadır. Bu bağlamda, madde x puanlayıcı ortak etkisi ve toplam varyansı açıklama oranının küçük olması, madde x puanlayıcı ortak etkisinden gelen farklarında küçük olduğunu, puanlayıcıların bireyleri belli bir maddeden diğerine aynı kararlılıkta puanladıkları biçiminde yorumlanabilir.

Top kontrol yetisi alt boyutunda artık varyansın birey x madde x puanlayıcı ortak etkisi varyans bileşeninin (0,20965) toplam varyansı açıklama oranını % 20,12 olduğu görülmektedir. Bu bağlamda birey x madde x puanlayıcı ortak etkisi ve/veya tesadüfi hataların oldukça büyük olduğunu göstermektedir.

Futbol Yetilerine İlişkin Dereceleme Ölçeğini puanlamada Top sürme yetisi alt boyutunda çok değişkenli analiz ile kestirilen en büyük varyans bileşeni, birey (b) temel etkisi için kestirilen varyans bileşeni (0,41830) ve toplam varyansın % 52,34'ünü açıklamaktadır. Bu bağlamda, top sürme yetisi, bireylerin ölçülen özellikleri boyutunda aralıklarındaki farklılıkların daha az ortaya konulabildiğinin bir göstergesidir.

Madde (m) temel etkisi için kestirilen varyans bileşeni (0,02125) toplam varyansın % 2,66'sını açıklamaktadır. Maddeler için kestirilen varyans bileşeni oranın sıfıra yakın olması, maddelerden kaynaklanan farklılıkların da küçük olduğu ve maddelerin yakın güçlükte olduğu şeklinde yorumlanabilir.

Puanlayıcı (p) temel etkisi için kestirilen varyans bileşeni (0,00533) ve toplam varyansın % 0,67'sini açıklamaktadır. Puanlayıcılar için kestirilen varyans bileşenin ve toplam varyansı açıklama oranının sıfıra yakın olması, puanlayıcıların

bütün bireyler için yaptıkları puanlamaları arasında farklılık bulunmadığı, puanlayıcıların bütün bireyleri birbirine yakın katılık/çömerlikte puanlar verdikleri şeklinde yorumlanmaktadır.

Birey x madde (bm) ortak etkisi (0,00883) toplam varyansın % 1,11'ini açıklamaktadır. Birey x madde ortak etkisinin bu boyutta varyans açıklama oranının küçük olduğu görülmektedir. Bu durum birey x madde ortak etkisinden kaynaklanan farklılıkların küçük olduğunu ve bireylerin bir maddeden diğerine durumlarının farklılaşmadığı şeklinde yorumlanmaktadır.

Birey x puanlayıcı (bp) ortak etkisi (0,03846) toplam varyansın % 4,81'ini açıklamaktadır. Birey x puanlayıcı ortak etkisi varyansının sıfıra yakın olması, belli bir puanlayıcının belli bir bireyi diğer puanlayıcılardan farklı puanlamadıkları şeklinde yorumlanmaktadır.

Madde x puanlayıcı (mp) ortak etkisi (0,01486) toplam varyansın % 1,86'sını açıklamaktadır. Bu bağlamda, madde x puanlayıcı ortak etkisi ve toplam varyansı açıklama oranının küçük olması, madde x puanlayıcı ortak etkisinden gelen farklarında küçük olduğunu, puanlayıcıların bireyleri belli bir maddeden diğerine aynı kararlılıkta puanladıkları biçiminde yorumlanabilir.

Artık varyansın birey x madde x puanlayıcı ortak etkisi varyans bileşeninin (0,29220) toplam varyansı açıklama oranını % 36,56 olduğu görülmektedir. Bu bağlamda birey x madde x puanlayıcı ortak etkisi ve/veya tesadüfi hataların oldukça büyük olduğu şeklinde yorumlanmaktadır.

Futbol Yetilerine İlişkin Dereceleme Ölçeğini puanlamada Şut yetisi alt boyutunda çok değişkenli analiz ile kestirilen en büyük varyans bileşeni, birey (b) temel etkisi için kestirilen varyans bileşeni (0,63693), toplam varyansın % 57,09'unu açıklamaktadır. Bireyler için kestirilen varyans bileşeninin toplam

varyans içinde en büyük payı alması, bireylerin ölçülen performansları bakımından aralarındaki farklılıkların ortaya konulabildiğinin bir göstergesidir.

Madde (m) temel etkisi için kestirilen varyans bileşeni (0,08106) toplam varyansın % 7,27'sini açıklamaktadır. Maddeler için kestirilen varyans bileşenin ve toplam varyansı açıklama oranının sıfıra yakın olması, maddelerden kaynaklanan farklılıkların da küçük olduğu ve maddelerin yakın güçlükte olduğu şeklinde yorumlanabilir.

Puanlayıcı (p) temel etkisi için kestirilen varyans bileşeni (0,04036) ve toplam varyansın % 3,62'sini açıklamaktadır. Puanlayıcı temel etkisinin sıfıra yakın olması, puanlayıcıların bütün bireyler için yaptıkları puanlamaları arasında farklılık bulunmadığı, puanlayıcıların bütün bireyleri birbirine yakın katılık/çömerlikte puanladıkları görülmektedir.

Birey x madde (bm) ortak etkisi (0,01633) toplam varyansın % 1,46'sını açıklamaktadır. Birey x madde ortak etkisinden kaynaklanan farklılıkların küçük olduğunu ve bireylerin bir görevden diğerine durumlarının farklılaşmadığı şeklinde yorumlanmaktadır.

Birey x puanlayıcı (bp) ortak etkisi (0,06323) toplam varyansın % 5,67'sini açıklamaktadır. Birey x puanlayıcı ortak etkisi varyansının sıfıra yakın olması, belli bir puanlayıcının belli bir bireyi diğer puanlayıcılardan farklı puanlamadıkları şeklinde yorumlanmaktadır.

Madde x puanlayıcı (mp) ortak etkisi (0,05752) toplam varyansın % 5,16'sını açıklamaktadır. Bu bağlamda, madde x puanlayıcı ortak etkisi ve toplam varyansı açıklama oranının küçük olması, madde x puanlayıcı ortak etkisinden gelen farklarında küçük olduğunu, puanlayıcıların bireyleri belli bir maddeden diğerine aynı kararlılıkta puanladıkları biçiminde yorumlanabilir.

Artık varyansın birey x madde x puanlayıcı ortak etkisi varyans bileşeninin (0,22020) toplam varyansı açıklama oranını % 19,74 olduğu görülmektedir. Bu bağlamda birey x madde x puanlayıcı ortak etkisi ve/veya tesadüfi hataların oldukça büyük olduğu şeklinde yorumlanmaktadır.

Alt boyutlara ait sonuçlar incelendiğinde bütün alt boyutlar için en yüksek varyans bileşeninin birey temel etkisi için kestirilen varyans bileşeni olduğu görülmektedir. Bunu bxm_{xp} ortak etkisi varyans bileşeni takip etmektedir. Bunların dışındaki temel ve ortak etkilere ait varyans bileşenlerinin ise sıfıra yakın olduğu görülmektedir.

Çok değişkenli G çalışması sonucunda elde edilen bu bulgu, Atılgan (2008)'in çalışmasındaki bazı bireylerin bir maddeden diğerine bağlı durumlarının farklılaşmasının az olduğu ve puanlayıcıların bireyleri bir görevden diğerine kararlı puanladıkları şeklindeki bulgularıyla tutarlılık göstermektedir. Ancak bazı puanlayıcıların bazı bireyleri diğerlerine göre daha katı ya da cömert puanladıkları şeklindeki bulgu ile tutarlı olmadığı görülmektedir.

c. Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin alt boyutları için madde sayılarının bir arttırılıp, bir azaltılması ve puanlayıcı sayılarının 3, 4 ve 5 olduğu koşullarda kestirilen G ve Phi katsayılarının derecesi nedir?

Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin alt boyutlara göre orijinal puanlayıcı ve madde sayıları ile puanlayıcı ve madde sayılarının bir arttırılıp bir azaltılması senaryolarına göre yapılan Karar çalışmaları sonucunda kestirilen G ve Phi (Φ) katsayıları Tablo 11'de verilmiştir.

Tablo 11. Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin Alt boyutları için K Çalışması Senaryolarına Göre Kestirilen G ve Phi Katsayıları

		Pas Yetisi		Top Kontrol Yetisi			Top Sürme Yetisi			Şut Çekme Yetisi		
n	nm	G	Phi	nm	G	Phi	nm	G	Phi	nm	G	Phi
3	16	0,83114	0,80154	10	0,69046	0,55146	11	0,74207	0,62239	15	0,79379	0,65415
3	17	0,83303	0,80405	11	0,70154	0,56807	12	0,75042	0,6319	16	0,79728	0,65875
3	18	0,83471	0,80630	12	0,71111	0,58288	13	0,75782	0,6406	17	0,80037	0,66285
4	16	0,86769	0,84070	10	0,75422	0,60783	11	0,80691	0,67169	15	0,85795	0,71407
4	17	0,86923	0,84292	11	0,76472	0,62495	12	0,81597	0,68268	16	0,86137	0,71927
4	18	0,87061	0,84492	12	0,77374	0,64012	13	0,82396	0,6927	17	0,86438	0,7239
5	16	0,89120	0,86608	10	0,79954	0,64895	11	0,85914	0,71337	15	0,90357	0,75826
5	17	0,89251	0,86811	11	0,80943	0,66623	12	0,86856	0,72544	16	0,90687	0,7639
5	18	0,89367	0,86992	12	0,81789	0,68148	13	0,87685	0,73640	17	0,90979	0,76893

Tablo 11'den elde edilen verilere göre Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin orijinal puanlayıcı ve madde sayılarının yanı sıra puanlayıcı sayılarının ve madde sayılarının bir artırılıp bir azaltılması senaryoları karşısında G ve Phi katsayılarının durumu ayrı alt başlıklar altında incelenmiştir.

c.1. G katsayılarının derecesi nedir?

Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin Pas Yetisi alt boyutunda orijinal puanlayıcı (4) ve madde (17) sayılarının yanı sıra puanlayıcı sayılarının bir artırılıp bir azaltılması ve madde sayılarının bir artırılıp bir azaltılması senaryoları karşısında G katsayılarının 0,83114 ile 0,89367 arasında değiştiği görülmektedir.

Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin Top kontrol yetisi alt boyutunda orijinal puanlayıcı (4) ve madde (11) sayılarının yanı sıra puanlayıcı sayılarının bir artırılıp bir azaltılması ve madde sayılarının bir artırılıp bir

azaltılması senaryoları karşısında G katsayılarının 0,69046 ile 0,81789 arasında değiştiği görülmektedir.

Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin Top Sürme Yetisi alt boyutunda orijinal puanlayıcı (4) ve madde (12) sayılarının yanı sıra puanlayıcı sayılarının bir arttırılıp bir azaltılması ve madde sayılarının bir arttırılıp bir azaltılması senaryoları karşısında G katsayılarının 0,74207 ile 0,87685 arasında değiştiği görülmektedir.

Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin Şut Yetisi alt boyutunda orijinal puanlayıcı (4) ve madde (16) sayılarının yanı sıra puanlayıcı sayılarının bir arttırılıp bir azaltılması ve madde sayılarının bir arttırılıp bir azaltılması senaryoları karşısında G katsayılarının 0,79379 ile 0,90979 arasında değiştiği görülmektedir.

c.2. Phi katsayılarının derecesi nedir?

Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin Pas Yetisi alt boyutunda orijinal puanlayıcı (4) ve madde (17) sayılarının yanı sıra puanlayıcı sayılarının bir arttırılıp bir azaltılması ve madde sayılarının bir arttırılıp bir azaltılması senaryoları karşısında Phi katsayılarının 0,80154 ile 0,86992 arasında değiştiği görülmektedir.

Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin Top kontrol yetisi alt boyutunda orijinal puanlayıcı (4) ve madde (11) sayılarının yanı sıra puanlayıcı sayılarının bir arttırılıp bir azaltılması ve madde sayılarının bir arttırılıp bir azaltılması senaryoları karşısında Phi katsayılarının 0,55146 ile 0,68148 arasında değiştiği görülmektedir.

Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin Top Sürme Yetisi alt boyutunda orijinal puanlayıcı (4) ve madde (12) sayılarının yanı sıra puanlayıcı sayılarının bir artırılıp bir azaltılması ve madde sayılarının bir artırılıp bir azaltılması senaryoları karşısında Phi katsayılarının 0,62239 ile 0,73640 arasında değiştiği görülmektedir.

Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin Şut Yetisi alt boyutunda orijinal puanlayıcı (4) ve madde (16) sayılarının yanı sıra puanlayıcı sayılarının bir artırılıp bir azaltılması ve madde sayılarının bir artırılıp bir azaltılması senaryoları karşısında Phi katsayılarının 0,65415 ile 0,76893 arasında değiştiği görülmektedir.

Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin puanlamasına ilişkin analiz sonuçlarına genel olarak bakıldığında, puanlama verilerine ait Phi katsayılarının artırılması için puanlayıcı sayılarının artırılmasının daha verimli olduğu sonucuna varılmıştır. Ayrıca her boyut için farklı puanlayıcı-madde senaryolarına göre kestirilen Phi katsayılarının G katsayılarından daha küçük olduğu bulunmuştur. Bununla birlikte, Phi katsayılarının artışına ve azalışına paralel olarak G katsayılarında arttığı ve azaldığı görülmektedir. Başka bir ifadeyle G katsayısının farklı puanlayıcı-madde senaryolarına göre artışı ya da azalışına paralel olarak Phi katsayıları da artmış ya da azalmıştır. Ancak boyutlara göre kestirilen G ve Phi katsayıları arasındaki farklılıkların değiştiği görülmektedir.

Futbol Yetilerine İlişkin Dereceleme Ölçeğine ait G ve Phi katsayılarından elde edilen bu bulgular, Johnson ve ark., (2000), çalışmasındaki puanlayıcı sayısının artırılmasının puanların güvenilirliğini artırdığı sonucu ile tutarlılık göstermektedir. Ancak Lee ve ark., (2002)'nin çalışmasındaki, puanlayıcı sayısının artırılmasının ek maliyet getirmesinden dolayı madde sayısının artırılmasının puanların güvenilirliğini artırdığı sonucu ile tutarlılık göstermemektedir.

3) Futbol Yetilerine İlişkin Dereceleme Ölçeği'nden elde edilen puanların Genellenabilirlik ve Klasik Test Kuramına dayalı güvenilirlik kestirimleri iç ölçütlere göre değerlendirildiğinde yeterli midir?

Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin puanlayıcılar tarafından puanlanmasından elde edilen verilere ilişkin KTK ve G Kuramı analizleri ile güvenilirlik katsayıları iç ölçütlere göre değerlendirildiğinde yeterli olup olmadığı sınınanmıştır.

Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin Puanlanmasına ait KTK ve G kuramına göre elde edilen Güvenirlik katsayıları Tablo 12'de verilmiştir.

Tablo 12. Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin Puanlanmasından Elde Edilen Verilere Göre KTK ve G kuramından Kestirilen Güvenirlik Katsayıları

Boyutlar	Boyutlar G katsayısı	G katsayısı	Boyutlar Phi katsayısı	Phi katsayısı	Boyutlar Cronbach Alfa	Cronbach Alfa	Boyutlar Kendall W	Kendall W
Pas Yetisi	0,87	0,82	0,84	0,77	0,91	0,86	0,77	0,78
Top Kontrol Yetisi	0,76		0,62		0,66		0,72	
Top Sürme Yetisi	0,82		0,68		0,68		0,80	
Şut Yetisi	0,86		0,72		0,90		0,83	

Tablo 12 incelendiğinde 56 maddeden oluşan Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin K çalışması ile kestirilen G katsayısının 0,82 olduğu görülmektedir. Dereceleme Ölçeğine alt boyutlar düzeyinde bakıldığında çok değişkenli K çalışması ile kestirilen G katsayılarının 0,76 ile 0,87 arasında değişen değerlere sahip olduğu görülmektedir.

Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin K çalışması ile kestirilen Phi katsayısının 0,77 olduğu görülmektedir. Ölçeğe alt boyutlar düzeyinde

bakıldığında çok deęişkenli K alıřması ile kestirilen Phi katsayılarının 0,62 ile 0,84 arasında deęişen deęerlere sahip olduęu grlmektedir.

Futbol Yetilerine İliřkin Dereceleme leęi'nin i tutarlılık katsayısı iin hesaplanan Cronbach alfa katsayısının 0,86 olduęu grlmektedir. leęin alt boyutlarına ait Cronbach alfa i tutarlılık katsayılarının 0,66 ile 0,91 arasında deęişen deęerlere sahip olduęu grlmektedir.

Futbol Yetilerine İliřkin Dereceleme leęi'nin puanlanmasında drt puanlayıcı arasındaki tutarlılıęı saptamak iin kullanılan Kendall'ın Uyuřum (Konkordans) katsayısı 0,78 olarak hesaplanmıřtır. leęe alt boyutlar dzeyinde bakıldığında Kendall'ın uyurřum katsayıları 0,72 ile 0,83 arasında deęişen deęerler elde edilmiřtir.

Bu bulgulara gre Futbol Yetilerine İliřkin Dereceleme leęi'nin puanlamasından elde edilen gvenirlik katsayıları i ltlere gre incelendięinde G katsayısı iin 0,80 ve Cronbach Alfa iin 0,70 genel kabullerine gre incelendięinde beklenen deęerlerinden yksek olduęu grlmektedir. Phi katsayısı ile Kendall W Gvenirlik katsayılarının 0,80 genel kabullerine gre incelendięinde beklenen deęerlerinden dřk olduęu grlmektedir.

Futbol Yetilerine İliřkin Dereceleme leęi'ne ait KTK ve G kuramı analizlerine gre elde edilen bu bulgular, Goodwin (2001), O'Brian ve ark., (2003) ve Yelboęa (2007)'nin alıřmalarındaki KTK ve G Kuramı gvenirlik katsayılarının i ltlere gre beklenen deęerlerinden yksek sonular rettięi řeklindeki bulgular ile tutarlılık gstermektedir.

BÖLÜM V

SONUÇ VE ÖNERİLER

Bu bölümde Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin puanlanmasından ortaya çıkan verilerin Genellenebilirlik ve Klasik Test Kuramı çalışmalarının analizleriyle elde edilen bulgulara dayalı olarak sonuçlara ve benzer çalışmalar için önerilere yer verilmiştir.

Sonuçlar

Futbol Yetilerine İlişkin Dereceleme Ölçeği ve alt boyutlarına göre uyuşum katsayıları incelendiğinde puanlayıcılar birbirine benzer puanlamalar yapmıştır.

Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin puanlanmasından elde edilen Karar çalışmasının orijinal madde ve puanlayıcı sayılarıyla kestirilen G katsayısı beklenen değer olan 0,80'den yüksek, ancak Phi katsayısının beklenen değer 0,80'den düşük olduğu bulunmuştur. K çalışması sonucunda G ve Phi katsayılarının her ikisinin de beklenen değer olan 0,80'in üzerine çıkması için, puanlayıcı sayısının bir arttırılması ($n_p=5$) gerektiği sonucuna ulaşılmıştır.

Alt boyutlar için G kuramının G çalışmasından değişkenlik kaynaklarına göre hesaplanan varyans yüzdeleri incelendiğinde;

a) Bireylerin alt boyutlarda ölçülen özellikleri bakımından aralarındaki farklılıkların ortaya konulabildiği, madde güçlüklerinin oldukça farklılaştığı ve puanlayıcıların bütün bireyler için benzer katılık/cömerlikte puanlama yaptıkları şeklinde yorumlanmaktadır.

b) Ortak etkileri için hesaplanan varyans yüzdeleri incelendiğinde bazı bireylerin bir maddeden diğerine bağıl durumlarının farklılaşmasının az olduğu, puanlayıcıların, bireyleri puanlamalarında birbirine bağımlı olmadığı, bireyleri bir

maddeden diğere kararlı puanladıkları ve ölçülemeyen hata kaynağının büyük olduğu şeklinde yorumlanmaktadır.

Alt boyutlara göre; puanlayıcı ve madde sayılarının birer artırılıp ve azaltılması ile yapılan alternatif K çalışmaları incelendiğinde;

a) Madde sayılarının birer azaltılması ile kestirilen G ve Phi katsayılarındaki azalma, puanlayıcı sayılarının birer azaltılması ile kestirilen G katsayılarındaki azalmadan daha düşük olduğu görülmektedir.

b) Madde sayılarının birer artırılması ile kestirilen G ve Phi katsayılarındaki artış, puanlayıcı sayılarının birer artırılması ile kestirilen G katsayısından daha düşük olduğu görülmektedir.

Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin alt boyutlarının puanlanmasına ait çok değişkenli model K çalışması sonucunda puanlayıcı ve madde sayısının birer artırılmasıyla, G katsayılarının, beklenen değerin (0,80) üstünde olduğu görülürken, Phi katsayıları için "Pas yetisi" dışındaki alt boyutlarda beklenen değerin altında olduğu görülmüştür.

Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin puanlamasına ilişkin verilere bakıldığında Phi katsayılarının artırılması için puanlayıcı sayılarının artırılmasının daha verimli olduğu sonucuna varılmıştır.

Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin puanlamasından elde edilen güvenilirlik katsayıları iç ölçütlere göre incelendiğinde G katsayısı ve Cronbach Alfa katsayıları beklenen değerlerinden yüksek olduğu görülmektedir. Phi katsayısı ile Kendall W Güvenirlik katsayılarının beklenen değerlerinden düşük olduğu görülmektedir.

Bu bağlamda, G kuramının, KTK'nın sınırlılıklarından olan farklı yöntemlerle elde edilen güvenilirlik katsayılarının birbirinden farklı olması durumuna alternatif olarak geliştirildiği ve potansiyel hata kaynaklarının birden fazla olması durumunda güvenilirlik katsayısının bir analizle bulunmasını sağladığı alanyazında bahsedilmektedir.

Sonuç olarak araştırmamızdaki potansiyel hata kaynaklarının fazla olduğu durumlara benzer çalışmalarda G kuramı, KTK'na alternatif oluşturmaktadır. Bununla birlikte, Futbol Yetilerine İlişkin Dereceleme Ölçeğinin ve alt boyutların güvenilirliğinin yüksek olması, ölçeğin güvenilir bir ölçek olarak değerlendirilebileceğini göstermektedir.

Öneriler

Bu bölümde, araştırmada ele alınan Futbol Yetilerine İlişkin Dereceleme Ölçeği uygulamasının iki farklı kurama göre elde edilen sonuçların doğrultusunda işevuruk öneriler sunulmuştur.

- Futbol Yetilerine İlişkin Dereceleme Ölçeği ile ilgili öneriler:

1- Ölçeğin, futbol yetilerindeki hazırbulunuşluluk ve gelişimini izlemeyi sağlayan bir araç olarak psikometrik özellikleri araştırılabilir. Buna göre, ölçeğin farklı yaş aralıklarında futbol yetilerine ilişkin normları çıkarılabilir.

2- Futbol Yetilerine İlişkin Dereceleme Ölçeği'nin farklı yaş gruplarında, her iki kurama göre psikometrik özellikleri sınanabilir..

3- Bu ölçekten alınan puanlarla futbol yetilerinde önemli olabilecek Pas, Top Kontrolü, Top sürme ve Şut yetileri arasındaki ilişkiler araştırılabilir.

- Klasik Test Kuramı ve Genellenebilirlik kuramı ile ilgili öneriler:

1- KTK ve G kuramının benzer sonuçlar ürettiği görülmüştür. Benzer çalışmalar için G kuramına dayalı olarak, farklı madde ve puanlayıcı sayıları ile alternatif analizler yapılarak optimum puanlayıcı ve madde sayılarına yönelik tahminler yapılabilir.

2- Farklı güvenirlik ve geçerlik katsayılarına sahip ölçeklerden elde edilen veriler ile aynı ölçeklerin G kuramından elde edilen güvenirlik katsayılarındaki tutarlılığa bakılarak, ölçmelerin psikometrik niteliklerindeki değişmeler araştırılabilir.

BÖLÜM VI

KAYNAKLAR

- Açıkada, C. ve Ergen, E. (1990) *Science and Sports*, Ankara: Büro-Tek Matbaacılık.
- Adıgüzel, M. (1985). "Madde Yapısının Test Güvenirliğine ve Geçerliğine Etkisi", Yayınlanmamış Doktora Tezi, Hacettepe Üniversitesi, Ankara.
- Aksoy, Ş. (1986) Introduction to Personel Administration. Yayınlanmamış Yüksek Lisans Tezi. Middle East Technical University, Ankara.
- Anastasi, A. (1988) *Psychological Testing*, New York: Macmillan Publishing Co. Inc.
- Ataay, İ.D. (1990). *İş Değerlendirme ve Başarı Değerlendirme Yöntemleri*. İşletme Fakültesi Yayınları, İstanbul.
- Atılğan, H. (2004). Genellenebilirlik Kuramı ve Çok Değişkenlik Kaynaklı Rasch Modelinin Karşılaştırılmasına İlişkin Bir Araştırma. Yayınlanmamış Doktora Tezi, Hacettepe Üniversitesi, Ankara.
- Atılğan, H. (2005a). Müzik Öğretmenliği Özel Yetenek Seçme Sınavının Çok-Yüzeyle Rasch Modeli ile Analizi (İnönü Üniversitesi Örneği). *Eurasian Journal of Educational Research*, 20, 62-73.
- Atılğan, H. (2005b). Genellenebilirlik Kuramı ve Puanlayıcılar Arası Güvenirlik için Örnek Bir Uygulama. *Eğitim Bilimleri ve Uygulama*, 4, (7) 95-108.
- Atılğan, H. ve Tezbaşaran A. A. (2005). Genellenebilirlik Kuramı Alternatif Karar Çalışmaları ile Senaryolar ve Gerçek Durumlar İçin Elde Edilen G ve Phi Katsayılarının Tutarlılığının İncelenmesi. *Eğitim Araştırmaları*, 18, 28 -40.
- 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.
- Balci, A. (2000) *Örgütsel Gelişme Kuram ve Uygulama*. Ankara: Pegem Yayıncılık.

- Balcı, A. (2001) *Sosyal Bilimlerde Araştırma: Yöntem, Yeti ve İlkeler*, Ankara: Pegem Yayıncılık.
- Barutçugil İ. (2002). *Performans Yönetimi*. İstanbul: Kariyer Yayınları.
- Başaran, İ. E. (1985) Örgütlerde İşgören Hizmetlerinin Yönetimi. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi*: Yayın No: 139, Ankara.
- Baykul, Y. (2000) *Eğitimde ve Psikolojide Ölçme: Klasik Test Teorisi ve Uygulaması*, Ankara: ÖSYM Yayınları.
- Baysal, M.E. (1999). Veri Zarflama Analizi ile Orta Öğretimde Performans Ölçme. Yayınlanmış Yüksek Lisans Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- Beach, D. S. (1980) *Personel: The Management of People at Work*. NewYork: Macmillan Publishing Co. Inc.
- Bompa, T. (1983), *Biomotor Abilities and the Methodology Their Development Theory and Methodology of Training*, Second Edition, Dubuque, Iowa: Kendall / Hunt Publishing Company .
- Brennan, R.L. (2001) *Generalizability Theory*. Springer- Verlag, New York.
- Bruning J.R ve Kintz BL. (1983) (Çev: Dönmez A.) *İstatistik*. Ankara: Gündoğan Yayınları.
- Carey LM. (1988) *Measuring and Evaluating School Learning*. London: Allyn and Bacon Inc.
- Carmines E.G. ve Zeller R.A. (1982) *Reliability and Validity Assessment*. 5th printing. Beverly Hills: Sage Publications Inc.
- 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. ve Gleser, G.C. (1972). Theory of Generalizability: A liberation of Reliability Theory. *British Journal of Statistical Psychology*. 11 (3), 435 – 439.

- Çıtak, G. G. (2007) Klasik Test ve Madde-Tepki Kuramlarına göre Çoktan Seçmeli Testlerde Farklı Puanlama Yöntemlerinin Karşılaştırılması. Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Dağdeviren, M (2005). Performans Değerlendirme sürecinin çok ölçütlü karar verme Yöntemleri ile Bütünleşik Modellenmesi. Yayınlanmamış Yüksek Lisans Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- Darst, P.W., Zakrajsek, D.B. ve Mancini, V. H. (1999) *Analyzing Physical Education and Sport Instruction*. Illinois: Human Kinetics Boks.
- Dawson B, ve Trapp R. G. (2001) *Basic and Clinical Biostatistics*. 3rd edition. New York: Lange Medical Books / McGraw International Editions.
- Demirtaşlı, N. Ç., (1995) Rasch Modelinin Raven Standart İlerlemeli Matrisler Testine Uygulanması ve Klasik Test Kuramı ile karşılaştırılması. Yayınlanmamış Doktora Tezi, Ankara Üniversitesi Sosyal Bilimler Enstitüsü Eğitimde Psikolojik Hizmetler Anabilim dalı, Ankara.
- Dimitrov, M. D. (2002) Reliability: Arguments for Multiple Perspectives and Potential Problems with Generalization Across Studies. *Educational and Psychological Measurement*, 62; 783-801.
- Eklblom, B. (1994), *Handbook of Sport Medicine ve Science Football (Soccer)*, IOC
- Erdoğan, İ. (1991). *İşletmelerde Personel Seçimi ve Başarı Değerlendirme Yetileri*. İstanbul: İşletme İktisadi Enstitüsü Yayınları.
- Erkuş, A. (2003) Psikometri Üzerine Yazılar. *Türk Psikologlar Derneği*. Yayın No: 24.
- 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.
- Ferah, A. (1992) *Futbol, Eğitim, Öğretim*. Ankara: Neyir Matbaası.
- Gay LR. (1985) *Educational Evaluation and Measurement*. 2nd edition. London: A Bell ve Howell Company.
- Goodwin, L. D. (2001). Interrater Agreement and Reliability, *Measurement in Physical education and Exercise Science*, 5 (1), 13-14.

- Gronlund, N.E. (1976). *Measurement and Evaluation Teaching*. New York: The Macmillan Publication, Inc.
- Gürsakar N. (2001) *Bilgisayar Uygulamalı İstatistik-I*. İstanbul: Alfa Yayınları.
- Gürses, Ç. ve Olgun, P. (1989) *Sportif Yetenek Araştırma Metodu*, İstanbul: Türk Spor Vakfı, Araştırma No: 1.
- Haladyna, T. M. (1999). *Developing and validating multiple-choice test items*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hambleton, R. K. ve Plake, B. S. (1995) Using an Extended Angoff Procedure to set Standart on Complex Peormance Assessment. *Applied Measurement in Education* 8, 41- 55.
- Hambelton R.K. ve Swaminathan H. (1985) *Item Response Theory: Principles and Application*, Kluwer: Nijhoff Publishing.
- Helms, L. S. (1999). Basic concepts in classical test theory: Tests aren't reliable, the nature of alpha, and reliability generalization as meta-analytic method. (*ERIC Document Reproduction Service No. ED 427 083*).
- Helvacı, M. A. (2002) Performans Yönetimi Sürecinde Performans Değerlendirmenin Önemi Ankara Üniversitesi. *Eğitim Bilimleri Fakültesi Dergisi*. 35(1-2) 155-169.
- Herbin, R. (1976) , *Soccer*, London: Robert Laffont Publications.
- John, A. R. and Malcom, S. J. (1991), *Select Soccer Drills*, Canada: Human Kinetics Publishers.
- 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
- Kıroğlu, G. (2001) Uygulamalı Parametrik Olmayan İstatistiksel Yöntemler, Ankara: Paymaş Yayınevi.
- Lee, G. ve Frisbie, D. A. (1999) Estimating Reliability Under a Generalizability Theory Model for Test Score Composed of Testlets. *Applied Measurement in Education*, 12 (3). 237-255.
- Lee, Y., Kantor, R. ve Mollaun, P. (2002) Score Dependability of the Writing and Speaking section of New TOEFL. *Educational Testing Service*.

- Lord, F. M. ve Novick, M. R. (1968) *Statistical Theory of Mental Test Test Scores*. New Jersey: Addison-Wesley. Co.
- Luxbacher, J.A. (1991) *Soccer Steps to Success*, Illinois: Leisure Press Campaign.
- Mehrens, W.A. (1992) Using Performance Assessment for Acoountability Purposes. *Educational Measurement*. 11, (4) 3-9.
- Morrow, J. R., Jackson, A.W., Disch, J. G.ve Mood, D.P. (2005) *Measurement and Evaluation in Human Performance*. United States: Human Kinetics.
- Murphy, K.R. ve Cleveland, J. N. (1995) *Understanding Performance Appraisal: Social, Organizational and Goal Perspectives*. Thousand Oaks, California: Sage Publication, Inc.
- Nunnally, J. ve Bernstein, I. (1994) *Psychometric Theory*. New York: McGraw Hill.
- O'Malley, J. M., ve Pierce, V. D. (1996) *Autentic Assessment*. Boston : Addison-Wesley Publishing Company Inc.
- Osburn, H. G. (2000). Coefficient alpha and related internal consistency reliability coefficients. *Psychological Methods*, 5, 343–355.
- Özçelik D.A. (1981) *Okullarda Ölçme ve Değerlendirme*. Ankara: ÜSYM-Eğitim Yayınları.
- Özdamar K. (2002) *Paket Programlarla İstatistiksel Veri Analizi-1.4*. Baskı. Eskişehir: Kaan Kitabevi.
- Özen, Y., Gülaçtı, F. ve Kandemir, M. (2006) Eğitim Bilimleri Araştırmalarında Geçerlik ve Güvenirlik Sorunsalı. *Erzincan Eğitim Fakültesi Dergisi*. 8(1). 69-89.
- Özgüven, İ. E. (1999). *Psikolojik Testler*. PDREM yayınları, Ankara, 1994.
- Özkara, A. (2004) *Futbolda Testler ve Özel çalışmalar*, Ankara: Kuşçu etiket ve matbaacılık.
- Reilly, T. (1996), *Motion Analysis ve Physiological Demands, Science ve Soccer*, London: E ve FN Spon.
- Sencer, M, ve Sencer Y. (1978) *Toplumsal Araştırmalarda Yöntembilim*. Ankara: Doğan Basımevi.

- Shavelson, R.J. ve Webb, M. N. (1991) *Generalizability Theory A Prime*. California: Sage Publication, Inc.
- Şencan, H. (2005) *Sosyal ve Davranışsal Ölçmelerde Güvenirlilik ve Geçerlik*. Ankara: Sözkese Matbaacılık.
- 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.
- Tavşancıl, E. (2002) *Tutumların Ölçülmesi ve SPSS ile Veri analizi*. Ankara: Nobel Yayın Dağıtım Ltd. Şti.
- Tekin H. (1977) *Eğitimde Ölçme ve Değerlendirme*. Ankara: Mars Matbaası.
- Thorndike R.M., Cunningham G.K., Thorndike R.L. ve Hagen E.P. (1991) *Measurement and Evaluation in Psychology and Education*. 5th edition. Macmillian Publishing Co.
- Tinsley, H.E.A., ve Weiss, D. J. (1975). Interrater reliability and agreement of subjective judgments. *Journal of Counseling Psychology*, 22.
- Traub R.E. (1994) *Reliability for the Social Sciences*. London: Sage Publications.
- 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.
- Turgut, F. ve Baykul, Y. (1992), *Ölçekleme Yetileri*, Ankara: ÖSYM Yayınları,
- Turgut F. M. (1993) *Eğitimde Ölçme ve Değerlendirme Metotları*. 9 baskı. Ankara: Saydam Matbaacılık.
- Turgut, F.M. (1997) *Eğitimde Ölçme ve Değerlendirme Metotları*. 10. baskı. Ankara: Yargıcı Matbaası
- Türker, U. A. (1998). *Yöneticinin El Kitabı: İnsan Kaynakları Yönetimi*. İstanbul: Türkmen Yayınevi.
- Walsh, W. B., ve Betz, N. E. (1990). *Tests and assessment*. Englewood Cliffs, NJ: Prentice Hall.
- Welk, J. G. (2002) *Physical Activity Assessments for Health- Related Research*. United States: Human Kinetics Publishers Inc.

- Worthington, E. (1974), *Teaching Soccer Skill*, Edinburg: Lepus Books, pp.73-80,
- Yelboğa, A., (2007) Klasik Test ve Genellenabilirlik Kuramına Göre Güvenirliğin Bir İş Performansı Ölçeği Üzerinde İncelenmesi Yayınlanmamış Doktora Tezi, Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Yıldırım, C. (1999). *Eğitimde Ölçme ve Değerlendirme*. ÖSYM Yayınları, Ankara.
- Yılmaz, S. (1992) *Futbol, Yeti*, Ankara: EHA Ajans.
- Yin, Y. ve Shavelson, R. J. (2004). Generalizability Analysis For Concept Mapping Assessment Of Students' Science Achievement. *Paper presented at the annual meeting of the AERA*, San Diego, CA.
- Yurdugül, H. (2006) Paralel, Eşdeğer ve Konjenerik Ölçmelerde Güvenirlik Katsayılarının Karşılaştırılması. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 39(1)15-37.

Ek 1. Futbol Yetilerine İlişkin Dereceleme Ölçeği

Futbol Yetilerine İlişkin Dereceleme Ölçeği

YÖNERGE: Aşağıda Futbolcuların futbol yetilerine (Pas, Top Kontrolü, Top sürme ve Şut) ait yeti bildiren bir grup ifade verilmiştir. Ölçekte bulunan ifadelerin her biri için her futbolcuyu ayrı ayrı değerlendirerek puanlamanız beklenmektedir. Futbolcuların video kamera ile çekilmiş görüntüleriyle “Futbol Yetilerine İlişkin Dereceleme Ölçeği” üzerindeki yeti ifadelerinden “Görülmedi”, “Yetersiz”, “Yeterli”, “Oldukça Yeterli” ve “Mükemmel” seçeneklerinden uygun gördüğünüz birini X ile işaretleyerek değerlendiriniz. Sizden beklenen, her bir yeti ifadesi bakımından video kaydında görüntüsünü izleyeceğiniz futbolcuyu değerlendirmenizdir. Eğer ilgili ifade de belirtilen yetiyi futbolcuda hiç gözlemediyseniz, “Görülmedi” kategorisini işaretleyiniz. Eğer ilgili yeti bakımından, yeti gösterilir ancak hareketi tamamlaması bakımından **destek olmazsa** “Yetersiz” kategorisi, yetinin doğru yapılması durumunda “Yeterli” kategorisi, yetinin mükemmel yakın gösterilmesi durumunda “Oldukça Yeterli” kategorisi ve yetinin mükemmel gösterilmesi durumunda “Mükemmel” kategorisini işaretleyiniz. Her futbolcu yetileri 5'er kez uygulayacaktır. Bütün uygulamaları dikkatlice izleyip işaretsiz hiçbir ifade bırakmayınız. Ayırdığınız zaman ve gösterdiğiniz işbirliği için teşekkür ederim.

PAS YETİSİ	GÖRÜLMEDİ	YETERSİZ	YETERLİ	OLDUKÇA YETERLİ	MÜKEMMEL
	0	1	2	3	4
Pas Hazırlık Safhası	0	1	2	3	4
1. Hedefi karşısına alması					
2. Omuzların hedefe dönük olması					
3. Denge ayağının topun yanında olması					
4. Denge ayağının bükülmesi					
5. Pas ayağının içinin topu göstermesi					
6. Pas ayağının geriye savrulması					
7. Başın aşağı bakması ve durağan olması					
8. Topa odaklanma					
Pas Uygulama Safhası	0	1	2	3	4
9. Vücudun topun üzerinde olması					
10. Pas ayağının öne savrulması					
11. Pas ayağının sıkı olması					
12. Ayağın iç kısmının kullanılması					
13. Topun merkezi ile temas					
Pas Uygulama Devam Safhası	0	1	2	3	4
14. Ağırlığın öne aktarılması					
15. Hareketin topa aktarılması					
16. Pasın bitirilmesi					
17. Hedefe dönük pozisyonun devam ettirilmesi					

TOP KONTROL YETİSİ	GÖRÜLMEDİ	YETERSİZ	YETERLİ	OLDUKÇA YETERLİ	MÜKEMMEL
Top Kontrolü Hazırlık Safhası	0	1	2	3	4
18. Vücutun Topa dönük olması					
19. Topa doğru hareketlenme					
20. Kontrol ayağın iç kısmının topa dönük olması					
21. Kontrol ayağın sıkı tutulması					
22. Kontrol ayağın topla temas etmesi için düzeltilmesi					
23. Topa odaklanma					
Top Kontrolü Uygulama Safhası	0	1	2	3	4
24. Kontrol ayağın içi ile topa temas edilmesi					
25. Topun hızının azaltılması için kontrol ayağın çekilmesi					
26. Rakipten uzak bir yere topun kontrol edilmesi					
Top Kontrolü Uygulama Devam Safhası	0	1	2	3	4
27. Diğer hareketin yönüne doğru topun itilmesi					
28. Sahaya odaklama için başın yukarı kaldırılması					
TOP SÜRME YETİSİ	GÖRÜLMEDİ	YETERSİZ	YETERLİ	OLDUKÇA YETERLİ	MÜKEMMEL
Top Sürme Hazırlık Safhası	0	1	2	3	4
29. Dizlerin bükülmesi					
30. Öne eğilme					
31. Ağırlık merkezinin düşürülmesi					
32. Vücutun topun üzerinde olması					
33. Başın kaldırılması					
Top Sürme Uygulama Safhası	0	1	2	3	4
34. Topa odaklanma					
35. Vücut yönünün değiştirilmesi					
36. Ayağın uygun yüzeyi ile topun itilmesi					
37. Sürat ve yönü değiştirme					
Top Sürme Uygulama Devam Safhası	0	1	2	3	4
38. Kontrolün devam etmesi					
39. Rakipten uzaklaşma					
40. Sahaya odaklanma için başın kaldırılması					
ŞUT YETİSİ	GÖRÜLMEDİ	YETERSİZ	YETERLİ	OLDUKÇA YETERLİ	MÜKEMMEL
Şut Çekme Hazırlık Safhası	0	1	2	3	4
41. Hafif açı ile topa yaklaşma					
42. Denge ayağını topun yanına yerleştirme					
43. Denge ayağının dizini bükme					
44. Denge için kolların yanlara açılması					
45. Şut ayağının geri çekilmesi					
46. Şut ayağının düzeltilmesi					

47. Bařın ařađıda ve sabit olması					
48. Topa odaklanma					
řut Çekme Uygulama Safhası	0	1	2	3	4
49. Omuzlar hedefe dönük					
50. Vücut topun üzerinde					
51. řut ayađın öne kapanması					
52. Ayađın üstü ile topun merkezine temas					
řut Çekme Uygulama Devam Safhası	0	1	2	3	4
53. Hareketin öne aktarılması					
54. Denge ayađın topun üzerinden kaldırılması					
55. řut çekmenin tamamlanması					
56. Giden topun izlenmesi					

Puanlamalarınız bitmiřtir.