



Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü
Eđitim Bilimleri Anabilim Dalı
Eđitimde Ölçme ve Deđerlendirme Bilim Dalı

**BUZ PATENİNDE HAKEM DEĐERLENDİRMELERİNİN
GENELLENEBİLİRLİK KURAMI VE RASCH MODELİ İLE
İNCELENMESİ**

Nihan ARSAN

Doktora Tezi

Ankara, 2012

BUZ PATENİNDE HAKEM DEĞERLENDİRMELERİNİN GENELLENEBİLİRLİK
KURAMI VE RASCH MODELİ İLE İNCELENMESİ

Nihan ARSAN

Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü
Eğitim Bilimleri Anabilim Dalı
Eğitimde Ölçme ve Değerlendirme Bilim Dalı

Doktora Tezi

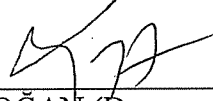
Ankara, 2012

KABUL VE ONAY

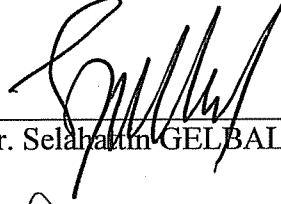
Nihan Arsan tarafından hazırlanan "Buz Pateninde Hakem Değerlendirmelerinin Genellenebilirlik Kuramı ve Rasch Modeli ile İncelenmesi" başlıklı bu çalışma, 27.12.2012 tarihinde yapılan savunma sınavı sonucunda başarılı bulunarak jürimiz tarafından Doktora Tezi olarak kabul edilmiştir.



Prof. Dr. Şener BÜYÜKÖZTÜRK (Başkan)



Doç. Dr. Nuri DOĞAN (Danışman)



Prof. Dr. Selahattin GELBAL



Doç. Dr. Hülya KELEÇİOĞLU



Yrd. Doç. Dr. Ziya KORUÇ

Yukarıdaki imzaların adı geçen öğretim üyelerine ait olduğunu onaylarım.

Prof. Dr. Yusuf ÇELİK

Enstitü Müdürü

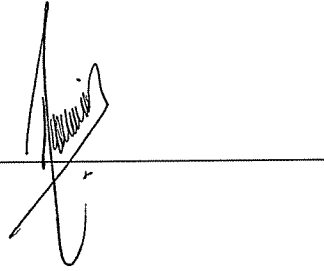
BİLDİRİM

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

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

27.12.2012

[Nihan Arsan]



CANIM ANNEME ve BABAMA

TEŞEKKÜR

Bu araştırmanın planlanmasından yazımına kadar her aşamasında bana yardımcı olan, değerli katkılarıyla beni yönlendiren, yapıcı önerileri ile beni yüreklendiren ve desteğini esirgemeyen saygıdeğer hocam ve tez danışmanım Doç. Dr. Nuri DOĞAN'a sonsuz teşekkürlerimi sunarım.

Doktora eğitimime başladığım ilk aşamadan itibaren desteklerini her zaman hissettiğim ve her sıkıntıda yardımlarını eksik etmeyen, sabırla dinleyen saygıdeğer hocalarım Prof. Dr. Selahattin Gelbal'a ve Hülya KELECİOĞLU'na çok teşekkür ederim.

Lisansüstü eğitime başladığım ilk günden beri yanımda olan, beni yönlendiren, yol gösteren, destekleyen ve sabırla dinleyip ümitsizliğe kapıldığım anlarda bana yardım elini uzatan değerli hocam Yrd. Doç. Dr. Ziya KORUÇ'a sonsuz saygı ve teşekkürlerimi sunarım.

Görüşleriyle ve eleştirileriyle çalışmama değerli katkılarda bulunan, bana güvenen ve bu süreci tamamlamamda desteklerini esirgemeyen sayın hocam Prof. Dr. Şener BÜYÜKÖZTÜRK'e saygılarımı sunar ve teşekkür ederim.

Ankara'da geçirdiğim süre boyunca bana anlayışlı davranan, sevgilerini esirgemeyen ve bu süreci anlamlı kılan, ümitsizliğe kapılmama izin vermeyerek beni ayakta tutan, onlar olmazsa yapamazdım dediğim çok değerli DOSTLARIMA ve sevgili öğrencilerime sabırlarından dolayı teşekkürler.

Bu hayattaki en büyük şansım ve zenginliğim olarak gördüğüm, her zaman yanımda olan ve her konuda olduğu gibi beni bu zorlu ve uzun süreçte de destekleyerek, yapabileceğime inandırarak, sabır ve anlayış gösteren, sevgilerini esirgemeyerek bu süreci tamamlamamı sağlayan ve hiç yalnız bırakmayan canım anneme, babama, ablama ve eşine sonsuz teşekkürlerimi sunuyorum. Ayrıca, yıllardır beni sabırla bekleyen, varlığıyla hayatımıza anlam, neşe ve mutluluk getiren, benim dünyalar tatlısı birtanecik yeğenim Eylül ÖZKAYALAR'a sonsuz sevgilerimi sunarım.

ÖZET

ARSAN, Nihan. *Buz Pateni Hakem Değerlendirmelerinin Genellenebilirlik Kuramı ve Rasch Modeli ile İncelenmesi*, Doktora Tezi, Ankara, 2012.

Bu araştırmada, Genellenebilirlik Kuramı ve Çok Değişkenlik Kaynaklı Rasch Modeli (ÇDKRM) ile buz pateni dünya şampiyonalarından elde edilen hakem değerlendirmelerine ilişkin gerçek veriler kullanılarak kestirilen parametrelerin karşılaştırılması amaçlanmıştır. Gerçek verilerde hangi yaklaşımın avantajlı olduğu, daha çok bilgi sağladığı ve iyi işleyip işlemediği ile bu yaklaşımların ürettiği bilgilerin birbirleriyle tutarlılıkları araştırılmıştır. Genellenebilirlik kuramına göre performans puanlamada çaprazlanmış desenler (sporcu_ihakem_jgörev) ve bu desenlerle yapılan karar çalışmalarının sonuçlarının karşılaştırılması ve karar çalışmalarının geçerliğinin test edilmesi amaçlanmıştır. Ayrıca, Rasch modeli kullanılarak, her bir değişkenlik kaynağı ayrıntılı olarak incelenmiştir.

Bu araştırmanın çalışma grubunu, 2006-2011 yılları arasında yapılan Dünya Buz Pateni Şampiyonalarında tekler ve çiftler kategorilerinde serbest programda yarışan toplam 397 sporcu ve bu yarışmalarda görev alan 189 hakem oluşturmaktadır. Araştırmada, 2006–2011 yılları arasındaki Dünya Buz Pateni Şampiyonalarında tekler ve çiftler kategorilerinde serbest programda, beş program bileşenine ilişkin hakem değerlendirmeleri kullanılmıştır.

Veriler genellenebilirlik kuramı ve ÇDKRM için kullanılan desenlere uygun olarak analiz edilmiştir. Genellenebilirlik kuramı hesaplamalarında EduG (Swiss Society for Research in Education Working Group, 2010) bilgisayar programı, ÇDKRM analizleri için FACETS (Linacre, 2007) programı kullanılmıştır.

Genellenebilirlik çalışması sonucunda, tüm yıllarda ve tüm gruplarda sporcu (s) ana etkisi için kestirilen varyans bileşeninin toplam varyansın büyük bir oranını (%76.6-%94.4) açıkladığı ve toplam varyans içerisinde ilk sırada yer aldığı görülmüştür. Çok değişkenlik kaynaklı Rasch ölçme modeline göre sporcu değişkenlik kaynağına ilişkin

ayırma indeksi güvenilirliğinin .99-1.0 değerler arasında değiştiği bulunmuştur. Bu da buz pateni performansının ölçülmesinde yer alan 5 görevde, bireyler arası farklılıkların ortaya çıkarılabildiği sonucuna varılmıştır.

Genellenebilirlik çalışmalarında hakemxsporcu etkileşimlerinin toplam varyansı açıklama oranını tüm yıllar ve tüm gruplar için %2.8-%14.4'ünü açıkladığı bulunurken, Rasch modeli ile yapılan yanlılık analizi sonucu bu oranlar %2.8-%13.7 bulunmuştur.

Buz pateni performansının ölçülmesinde hakem sayısının 12 olduğu yarışmalarda (2006-2008) G-katsayısı (.98 ve 1.00) ve Phi katsayısı (.98 ve 1.00) oldukça yüksektir. Karar çalışmasında hakem sayısının 9'a indirilmesi sonucunda G (.98 ve 1.00) ve Phi (.98 ve .99) katsayılarında düşüş meydana gelmemiş, bağıl ve hata varyanslarında da çok az bir artış meydana gelmiştir. 2009–2011 yıllarında elde edilen G (.98 ve .99) ve phi katsayılarının (.98 ve .99) yüksek olduğu görülmektedir. Karar çalışmaları açısından incelendiğinde ise 2009–2011 verilerinde hakem sayısının 12'ye çıkarılması artışa neden olmamaktadır. Bunun yanında, 2009–2011 yıllarında Dünya Buz Pateni Şampiyonalarında hakem sayısında azaltmaya gidilmesinin, genellenebilirlik ve güvenilirlik (ϕ) katsayılarını olumsuz etkilemediği sonucuna varılmıştır. Ayrıca, elde edilen bulgulara göre, karar çalışmaları geçerli sonuçlar vermektedir.

Genellenebilirlik kuramı ve ÇDKRM sundukları avantajlar bakımından, birbirini tamamlayıcı niteliktedir. Bu nedenle, değerlendirmelerin daha geniş kapsamlı (bireysel ve grup düzeyinde, etkileşimleri dikkate alarak) yapılması için her iki tekniğin kullanılması yarar sağlayabilir.

Anahtar Sözcükler

Buz pateni, Genellenebilirlik Kuramı, Rasch Modeli, karar çalışması, yanlılık.

ABSTRACT

ARSAN, Nihan. *Investigation of the Raters' Assessment in Ice Skating with Generalizability Theory and Rasch Measurement*, Ph.D. Dissertation, Ankara, 2012.

The purpose of this study is to compare Generalizability Theory and Many Facet Rasch Measurement Model based on estimated parameters from raters' assessment scores that obtained from Ice Skating World Championships. More advantageous, more informative and more operative theory has aimed to investigated in real (authentic) data. Also, consistency of information which generated from those two theories were investigated. In generalizability theory; crossed designs (skaterxjudgextask), decision studies in these designs were compared in this study. In this process, validity of decision studies has been examined. In addition to this, detailed information have acquired for each facet by using many facet Rasch measurement model.

397 ice skaters who performed in Ice Skating World Championships between 2006 and 2011 and 189 raters who judged them in these competitions were consisted research group. Raters' assessments which obtained from singles (men and women) and pairs competition in five program component were used as a research data. Data were analyzed based on designs that are appropriate for generalizability theory and many facet rasch measurement models. For generalizability theory analyses EduG (Swiss Society for Research in Education Working Group, 2010) and for many facet Rasch measurement model analyses FACETS programs were used.

As a result of generalizability analyses, variances due to skaters were too high. Skaters' variances were found between 76.6%-94.4% of the total variances for all years and groups. According to Rasch analysis, the reliability of skater ranges between .99 and 1.0. So, in the measurement of the ice skating performances, differences between skaters in five program components were identified.

Variances due to interactions between skatersxraters in all years and groups were found between 2.8%-14.4% of the total variances. In Rasch model, bias analyses were provided similar results (2.8%-13.7%).

According to generalizability analyses, the generalizability and reliability coefficients were between .98-1.00 in the 2006-2008 competitions. Decision studies were conducted for changing raters' number from 12 to 9. Decision studies' results showed that generalizability (.98-1.00) and reliability coefficients (.98-.99) were not decreasing. In addition to this, absolute and relative error variances were not increasing so much. Generalizability and reliability coefficients (.98-.99) were high for competitions between 2009 and 2011 in all groups. In these years, raters number were 9 and decision studies were conducted for 12 raters. As a result, increasing the number of raters were not change the generalizability and reliability coefficients. These results were indicating the validity of D-studies.

Both methods provide the information in different levels (group level statistics and individual level statistics). They had different advantageous and they complement each other. So, it is beneficial to use both methods in multi-faceted conditions.

Key Words

Ice skating, Generalizability Theory, Rach Model, decision study, bias.

İÇİNDEKİLER

KABUL VE ONAY	i
BİLDİRİM.....	ii
ADAMA.....	iii
TEŞEKKÜR.....	iv
ÖZET.....	v
ABSTRACT.....	vii
İÇİNDEKİLER	ix
TABLolar DİZİNİ	xii
ÖNSÖZ.....	xiv
BÖLÜM I	1
GİRİŞ	1
1.1. PERFORMANS DEĞERLENDİRME	2
1.2. PERFORMANS DEĞERLENDİRMENİN NİTELİĞİ	11
1.2.1. Geçerlik ve Güvenirlik.....	11
1.3. PUANLAYICI HATALARININ TÜRLERİ	15
1.3.1. Cömertlik/Katılık	16
1.3.2. Ranj Kısıtlanması (Ranj Daralması).....	17
1.3.3. Halo Etkisi	19
1.3.4. Doğruluk/Yanlışlık (Accuracy/Inaccuracy).....	20
1.3.5. Tutarsızlık	20
1.3.6. Mantıksal Hata.....	21
1.3.7. Yakınlık (Proximity Error)	21
1.3.8. Yenilik, Yakın Zamanda Meydana Gelme/ Öncelik (Recency /Primacy Error)	22

1.3.9. Sıra Etkisi (Order Effects)	22
1.3.10. Algı Farklılıkları (Perception Differences).....	22
1.3.11. Taşıma, Devam Etme (Carryover Effects)	23
1.3.12. Zıtlık (Contrast Error).....	23
1.4 ÖLÇME HATALARINI BELİRLEMEDE KULLANILAN YAKLAŞIMLAR..	23
1.4.1. Klasik Test Kuramı.....	24
1.4.2. Genellenebilirlik Kuramı	30
1.4.3. Madde Tepki Kuramı (MTK)	37
1.4.4. Genellenebilirlik Kuramı ve Çok Değişkenlik Kaynaklı Rasch Modelinin Karşılaştırılması	46
1.5. PROBLEM DURUMU	48
1.6. PROBLEM CÜMLESİ.....	50
1.7. ALT PROBLEMLER.....	50
1.8. SINIRLIKLAR.....	51
1.9. ARAŞTIRMANIN AMACI ve ÖNEMİ	51
1.10. TANIMLAR VE KISALTMALAR.....	52
1.11. İLGİLİ ARAŞTIRMALAR.....	52
1.11.1. Yurtiçinde Yapılan Çalışmalar	53
1.11.2. Yurtdışında Yapılan Çalışmalar.....	56
1.11.3. Buz Pateni ile İlgili Araştırmalar	61
BÖLÜM II.....	69
YÖNTEM.....	69
2.1 ARAŞTIRMANIN TÜRÜ	69
2.2 ÇALIŞMA GRUBU	69
2.3 ARAŞTIRMA VERİLERİ	69
2.4 VERİLERİ ÇÖZÜMLEME TEKNİKLERİ.....	72

BÖLÜM III	75
BULGULAR VE YORUM.....	75
BÖLÜM IV	109
SONUÇ ve ÖNERİLER.....	109
4.1 SONUÇLAR	109
4.2 ÖNERİLER	112
KAYNAKLAR	114
EKLER.....	129

TABLOLAR DİZİNİ

Tablo 1. Araştırma grubuna ve araştırma verilerine ilişkin bilgiler.....	71
Tablo 2. 2006-2011 yıllarındaki Dünya Buz Pateni Şampiyonası erkekler serbest program sonuçlarına ilişkin varyans bileşenleri ve toplam varyansı açıklama yüzdeleri.....	76
Tablo 3. 2006–2011 yıllarındaki Dünya Buz Pateni Şampiyonası kadınlar serbest program sonuçlarına ilişkin varyans bileşenleri ve toplam varyansı açıklama yüzdeleri.....	79
Tablo 4. 2006–2011 yıllarındaki Dünya Buz Pateni Şampiyonası çiftler serbest program sonuçlarına ilişkin varyans bileşenleri ve toplam varyansı açıklama yüzdeleri.....	82
Tablo 5. 2006–2008 yıllarında 12 hakem ile yapılan genellenebilirlik ve 9 hakem ile yapılan karar çalışmalarından elde edilen G ve Phi katsayıları ile hata varyansları.....	84
Tablo 6. 2009–2011 yıllarında 9 hakem ile yapılan genellenebilirlik ve 12 hakem ile yapılan karar çalışmalarından elde edilen G ve Phi katsayıları ile mutlak ve bağıl hata varyansları.....	85
Tablo 7. 2006–2008 Dünya Buz Pateni Şampiyonası serbest program değerlendirmelerine ilişkin ÇDKRM analizleri sonucunda elde edilen standart artık yüzdeleri, sporcu, görev ve hakem değişkenlik kaynaklarına göre uyumsuzluk yüzdeleri.....	87
Tablo 8. 2006–2008 Dünya Buz Pateni Şampiyonaları serbest program değerlendirmelerine ilişkin ÇDKRM değişkenlik kaynaklarına ait logit minimum/maksimum, logit, iç uyum ve dış uyum ortalama ve standart sapma değerleri.....	89
Tablo 9. 2006–2008 Dünya Buz Pateni Şampiyonası serbest programa ilişkin sporcular için ayırma oranı, güvenilirlik ve X^2 testleri.....	90

Tablo 10. 2006–2008 Dünya Buz Pateni Şampiyonası serbest programa ilişkin görevler için ayırma oranı, güvenilirlik ve X^2 testleri.....	92
Tablo 11. 2006–2008 Dünya Buz Pateni Şampiyonası serbest programa ilişkin hakemler için ayırma oranı, güvenilirlik ve X^2 testleri.....	94
Tablo 12. 2009–2011 Dünya Buz Pateni Şampiyonaları serbest program değerlendirmelerine ilişkin ÇDKRM analizleri sonucunda elde edilen standart artık yüzdeleri, sporcu, görev ve hakem değişkenlik kaynaklarına göre uyumsuzluk yüzdeleri.....	96
Tablo 13. 2009–2011 Dünya Buz Pateni Şampiyonaları serbest program değerlendirmelerine ilişkin ÇDKRM değişkenlik kaynaklarına ait logit minimum/maksimum, logit, iç uyum ve dış uyum ortalama ve standart sapma değerleri.....	98
Tablo 14. 2009–2011 Dünya Buz Pateni Şampiyonası serbest programa ilişkin sporcular için ayırma oranı, güvenilirlik ve X^2 testleri.....	99
Tablo 15. 2009–2011 Dünya Buz Pateni Şampiyonası serbest programa ilişkin görevler için ayırma oranı, güvenilirlik ve X^2 testleri.....	101
Tablo 16. 2009–2011 Dünya Buz Pateni Şampiyonası serbest programa ilişkin hakemler için ayırma oranı, güvenilirlik ve X^2 testleri.....	103

ÖNSÖZ

“Buz Pateni Hakem Deęerlendirmelerinin Genellenabilirlik Kuramı ve Rasch Modeli ile İncelenmesi” isimli doktora tezi (Proje No: 012 T06 707 001), Hacettepe Üniversitesi, Bilimsel Arařtırmalar Birimi tarafından desteklenmiştir.

BÖLÜM I

GİRİŞ

Bu bölümde problem durumu, problem cümlesi, alt problemler, sınırlıklar, sayıtlar, araştırmanın amacı, önemi ve ilgili araştırmalar yer almaktadır.

Performansın geçerli, güvenilir ve adil bir şekilde ölçülmesi, ölçmenin temelinde yer almaktadır. Belirli kararlar vermek için birçok alanda ölçme yapılmaktadır. Hakemler, jimnastikçilerin ve buz patencilerin performansı, degüstatörler şarabın kalitesi, işverenler başvuranların işe uygunluğu, zoologlar hayvanların taşıdıkları özelliklere göre soyları ve eğitimciler ise öğrencilerin başarıları hakkında kararlar vermede ölçmeden yararlanmaktadırlar. Bu ölçümler sonucunda önemli kararlar verileceği zaman, bunların doğru ve adil olarak yapılması oldukça önemlidir. Sonuç olarak, puanlayıcı seçimi, eğitimi ve ölçümlerin kontrolünün yapılması, puanlayıcı hatalarının veya puanlama yanlışlığının etkisinin en aza indirilmesinde önemli bir yere sahiptir.

Spor ortamı psikomotor davranışların ölçüldüğü ve değerlendirildiği bir alandır. Birey, hedef alınan yetiyi uygular; puanlayıcı, uygulama sırasında bireyin hareketlerini gözler ve puan verir. Bu yöntem, aslında, davranışın gözlenip ölçülmesidir (Turgut, 1993).

Buz pateni gibi spor yarışmalarında da sporcunun performansı, hakem jürisi tarafından gözlenip program bileşenleri (kayma becerileri, geçişler, performans, koreografi ve yorumlama) ve sergilenen hareketler açısından değerlendirilmektedir. Puanlamada yansızlık önemlidir ve hakemler arası güvenilirliği gerektirmektedir.

Güvenirlilik, ölçme sonuçlarının tesadüfi hatalardan arınık olma derecesi olarak tanımlanmaktadır (Baykul, 2000). Güvenirlilik, aynı bireyler üzerinde yapılan bir niteliğe ait ölçümlerin benzer şartlarda tekrar elde edilebilirliğidir (Crocker ve Algina, 1986).

Performansın ölçülmesinde birçok hata kaynağı (ölçme araçları, ölçme yöntemi, ölçme işlemi yapan kişi, ölçmenin yapıldığı ortam, bireyin kişisel özellikleri) vardır (Nunally ve Bernstein, 1994). Buz pateninde, ölçmeye karışan hataların en önemli nedenlerinden

biri de puanlamanın hakemlerin kararlarına bağılı olarak yapılmasıdır. Burada hakem en önemli hata kaynağı olarak görülse de sporcuların sergiledikleri görevler ve performans bileşenleri de güvenilirliği etkileyen hata kaynaklarıdır. Bundan dolayı, güvenilirliği hesaplarken sadece hakemleri (hakemler içi veya hakemler arası) dikkate almak bize tam bir güvenilirlik tahmini vermeyecektir. Güvenirliği hesaplarken diğere hata kaynaklarını işleme alarak, bu kaynaklar arasındaki etkileşimleri göz önünde bulundurmanız gerekmektedir.

Güvenirliği değerlendirmede birçok yaklaşım veya teknik kullanılmaktadır. Bunlar; hakemler arasındaki uyum oranı veya yüzdesi, Cohen'in kappası, "ağırlıklandırılmış kapp", Kendall W uyum katsayısı, pearson korelasyon katsayısı, ortalamaların karşılaştırılması (t- testi veya tek yönlü varyans analizi), genellenebilirlik kuramı (Crocker ve Algina, 1986) ve çok değişkenlik kaynaklı Rasch modeli (Linacre, 1989) gibi tekniklerdir.

1.1. PERFORMANS DEĞERLENDİRME

Performans, bir görevin ne derecede ortaya koyulduğunun değerlendirilmesidir. Performansı değerlendirme ve hakkında karar vermek için, bilgi toplamak ve ölçme yapmak gerektiğinden ölçme ve değerlendirme burada önem kazanmaktadır. Ölçme ve değerlendirme sürecinde altı aşama dikkat çekmektedir. Bunlar; ölçülecek özelliklerin seçimi, ölçme için uygun yöntemin seçimi, veri toplama, toplanan verinin analiz edilmesi, karar verme ve bu kararların uygulanmasıdır (Nande ve Vali, 2010).

Performans değerlendirme, belirli ölçütler dâhilinde bireylerin performanslarını sergilediği, gözlenebilir görevlere dayanan değerlendirmelerdir (Spears, 2008). Russell ve Airasian (2012) performans değerlendirmeyi bireylerin ürün veya performans ortaya koyarak bir beceriyi ya da bilgiyi sergilemelerini gerektiren değerlendirme şekli olarak tanımlamıştır. Bireylerin bildiklerini ve bunu gerçek durumlarda yapabileceklerini sergilemelerine olanak sağlamaktadır. Performans değerlendirme psikometride gittikçe artan bir şekilde dikkat çekmektedir. Performans değerlendirme kavramı, bireyin becerilerini gözlemlemeye ve değerlendirmeye dayanmaktadır. Bunlara örnek; açık

uçlu sorular, aktif katılımı gerektiren görevler, bilgisayar simülasyonları, becerileri sergileme ve portfolyolardır (Linn, Baker ve Dunbar, 1991). Bu tür değerlendirmelerin amacı performansın kendisi veya performansın sonucu ve ürünüdür. Bireyin bir aracı kullanma yeterliğini değerlendirmek için bireyi yol testine almak performansa örnek olarak verilebilir. Diğer yandan, bireyin dile ilişkin yeterliği yazdığı kompozisyon ile değerlendirilebilir. Kısacası, performans değerlendirme; performans ve ürünün değerlendirilmesidir (Fitzpatrick ve Morrison, 1971).

Değerlendirme öğretme-öğrenme sürecinin önemli ve tamamlayıcı kısımlarından biridir. Performans değerlendirme, bireyin davranışı veya sorunun cevabını oluşturmasını veya yaratmasını ve bunu yaparken de birden fazla beceriyi, bilgiyi ve üst düzey düşünme becerilerini kullanmasını gerektirmektedir. Performans testleri, bireyin gerçekte (gerçek hayatta) meydana gelen bazı “ölçüt” durumları temsilen görevi ortaya koyma kapasitesini değerlendirmemize yardımcı olmaktadır (Linn, 1993). Performansı sergileyen kişi, görevi tamamlarken “bildiğinin/nasıl bildiğinin” yanında bunu “göstermeli”dir de (Iramaneerat ve Yudkowsky, 2007).

Performans değerlendirme, deneyimlerin sergilenmesine olanak sağlayan değerli bir alternatif olarak görülebilir. Bireylerin kendi seçimleri doğrultusunda, yapılandırdıkları cevaplar veya davranışların ne kadar iyi ortaya koyulduğunu değerlendirmektir. Eğitim açısından ele alındığında, öğrencilerin sunulan seçenekler arasından doğru cevabı seçmeleri yerine kendi cevaplarını yapılandırmalarını veya bir görevi sergilemelerini gerektirmektedir. Hedeflenen bilgi veya becerilerin gerçek durumlarda ya da gerçek durumlara benzer koşullarda veya ortamlarda uygulanması sonucunda öğrencilerin öğrenme düzeylerini ölçmeyi hedeflemektedir (AERA, APA ve NCME, 1999).

Performansa dayalı değerlendirmede birey performansı sergilerken, hakem tarafından gözlenmekte ve bu ortamda değerlendirilmektedir. Böylece hakem, bireyin hareketinden yorumlanabilir bir sonuç çıkarmakta ve bunu puanlamaktadır (McNamara, 1996). Performans değerlendirme sürecinde; test, bireyin performansını ortaya koymasını sağlamak için kullanılmaktadır. Performans ortaya koyulduktan sonra hakem

dereceleme ölçeğini veya ölçme aracını kullanarak, performansı değerlendirmekte ve bir puan vermektedir (Landy ve Farr, 1980).

Performans testleri ayrıca bilişsel becerilerin değerlendirilmesine de yardımcı olmaktadır. Performans değerlendirme, beceriyi veya davranışı gözlemlemeden, yeterliğin değerlendirilemeyeceği durumlarda da kullanılmaktadır. Dil becerilerini, araç kullanımını, sanatı, müziği, sporsal becerileri ve fiziksel uygunluğu değerlendirmede oldukça popülerdir ve yaygın olarak kullanılmaktadır (Resnick ve Resnick, 1996). Eğitim ortamlarında da artan bir şekilde, öğrenme çıktılarını değerlendirmek amacı ile yaygın olarak kullanılmaktadır. Kısa veya geniş açıklamalar gerektiren yapılandırılmış görevler, materyallerin veya araçların aktif olarak kullanımını gerektiren görevler bunlara örnek olabilir (Bond, Braskamp ve Roeber, 1996).

Bilişsel alanda olduğu gibi psikomotor alanda da değerlendirme önemli bir yere sahiptir. Spor, psikomotor becerilerin öne çıktığı bir alandır ve spor performansı motor performans ile ilişkili öğelere dayanmaktadır. Değerlendirme sürecinin en kritik aşamalarından biri hangi bilginin kullanılacağına karar verilmesidir (Nadeau, Richard ve Godbout, 2008).

Performans değerlendirme ile ilgili birçok çalışma genellikle sonuçlara odaklanmaktaydı (Gould, Petlichkoff ve Weinberg, 1984). Daha sonraları spor psikolojisi çalışmalarının (Gould, Petlichkoff, Simons ve Vevera, 1987) da katkıları ile sadece sonuca odaklanmanın gerçek performansı ortaya koymadığı, sonuçlara dayanan performans değerlendirmenin standart bir ölçüm olmadığı düşünülmüştür (Rees, Hardy ve Ingledew, 2000). Çünkü rakibin standartları zaman içerisinde değişmekte ve geçerli karşılaştırmalar zayıf temellere dayanmaktadır. Weinberg (1990) sadece sonuca odaklanmanın gerçek performans kalitesini sakladığını ve bireyin ne kadar iyi performans ortaya koyduğunu tam olarak yansıtmadığını savunmaktadır.

Sadece sonuçlara odaklanmak yerine sürecin de dikkate alınması gerektiği vurgulanmaktadır. Süreç ölçümleri farklı spor dallarındaki veya farklı performanslardaki görev karmaşıklığını daha iyi yansıtacaktır. Bu alanda çalışan

arařtırmacılarından Godbout (1990) dört genel kategoriye ieren iki boyutlu bir model önermiřtir. Bu modele gre, bir yandan oyuncun performansının teknik (hareket) veya taktik (karar verme) zellikleri dikkate alınırken, diđer yandan oyuna iliřkin hareketlerin sonularına (rn) veya bu hareketleri ortaya koyma řekline (sre) odaklanılmaktadır.

Sporda performansı lmede ma veya msabaka sırasında elde edilen verilerin yanında bir takım nicel ve nitel veriler de kullanılmaktadır. Godbout (1990) ile Grehaigne ve Godbout (1998), sporda performansı lmede kullanılacak stratejileri dört bařlıkta ele almıřlardır. Bunlar; standart testler, oyundan elde edilen istatistikler, standart ortamlarda puanlama leklerinin kullanılması ve dođal ortamlarda puanlama leklerinin kullanılmasıdır.

Standart motor beceri testleri, bireyin performansına iliřkin teknik rnler ile ilgili bilgi sađlamaktadır; belirli bir zaman ierisinde verilen pas sayısı; belirli bir deneme sayısı sonucunda gol veya basket olan řut sayısı, nceden belirlenmiř bir beceri rntsn belirli bir sre ierisinde dođru bir řekilde tamamlama. Gzlem hatalarının dıřında, bu testlerde puanlayıcılar arası gvenirlik olduka yksektir. Standart test kořullarında performans istikrarlıdır. Oyunun taktiksel bileřenleri dikkate alınmadıđı srece, bu her ne kadar zor olsa da, testler geerlidir. Gerekten de, test kořulları ne kadar standardize olursa, verilecek kararlar da o kadar hatasız olacaktır (Tritschler, 2000).

Standart kořullardaki puanlama leklerinin, performansın srele ilgili zelliklerini deđerlendirmede kullanımı son zamanlarda artmıřtır. Pinheiro (1994) yaptđı alıřmalarda, motor becerilerin niteliđini (řekil olarak) deđerlendirmede dereceleme (puanlama) leklerinin kullanımını aıklamıřtır. Bu lme senaryolarında, gzlemci (genellikle birden fazla gzlemci bulunmakta), nceden belirlenmiř bir becerinin gsterimini izledikten sonra bu dereceleme (puanlama) leklerinden birini kullanmakta ve performansın niteliđine iliřkin yargıda bulunmaktadır. Genellikle, lme aracındaki đeler srele ilgili (motor becerinin ortaya koyulma řekli) olmasına rađmen, gzlemde kullanılan ara motor becerinin sonucu ile ilgili maddeleri de iermektedir. Bu, sonucun aık bir řekilde tanımlanabildiđi ve standart ortamda sergilenebilecek bir durum

olduğunda söz konusu olabilir. Örneğin; pas atma, gol atma vb. Burada seçilen beceriye ilişkin hem süreç hem de sonuç bilgisi elde edildiği halde, bu stratejinin de geçerliği sınırlı olmaktadır. Burada yine, oyunun taktiksel bileşenleri dikkate alınmamaktadır.

Dereceleme (puanlama) ölçekleri oyun (doğal) ortamlarında kullanılabilir. Kuramsal olarak bu en geçerli yöntemlerden biridir. Performansın tüm öğelerini dikkate aldığından dolayı, canlı performans koşullarında tek bir gözlemci/hakem yeterli olmayabilir. Doğal ortamda performansı değerlendirmenin en önemli getirisi, performansın taktiksel özelliklerini değerlendirmektir. Ölçme daha çok süreç yönelimlidir ve gözlemciler hareketlerin etkililiğini ve uygunluğunu daha uygun bir şekilde değerlendirebilmektedirler (Nadeau, Richard ve Godbout, 2008).

Derecelendirme ölçekleri, yaygın olarak kullanılmalarına rağmen sağladıkları bilgilere ve kullanımlarına şüphe ile yaklaşılmakta ve eleştirilmektedirler. Birçok kez elden geçirilmelerine rağmen, bu düzeltmeler ve yenilemeler ne kadar iyi düzeyde yapılırsa yapılsın, dereceleme ölçeklerinden elde edilen verilerin özneliği, yanlılığı ve en kötüsü de amaçlı olarak çarpıtılmış olması konusunda korkular doğurmaktadırlar. Bu şüpheler karışığında, dereceleme ölçeklerinden elde edilen verilere güvenen ve bu verilere dayalı çıkarımlar yapan psikologlar ile diğer profesyonellerin, ölçümlerin doğruluğunu ve genel psikometrik özelliklerini değerlendirmek için teknikler ve işlem yolu geliştirmeye çok fazla zaman ve çaba harcamasına şaşırılmaması gerekir. Stratejilerin birçoğu ölçme hatalarını belirlemek için tasarlanmıştır. Ölçme kalitesine ilişkin indeksler veya ölçütler geliştirilerek ölçümlerin “çok öznel” olduğuna ilişkin algıyı azaltma çalışmalarının sonucunda da karmaşa meydana gelmiştir. İlk olarak, puanlama kalitesinin ölçütleri ile ilgili kavramsal tanımlardan birkaç tanesine ilişkin fikir birliği çok azdır. İkinci olarak, bu ölçütler için işlevsel tanımlara (istatistiksel indeksler) ilişkin fikir birliği de azdır. Üçüncü olarak, farklı araştırmacılar farklı araştırma desenleri kullanmakta veya farklı veri toplama işlemleri uygulamaktadırlar, bu da verileri bir araya getirme kapasitesini sınırlandırmakta ve puanlamanın niteliğine ilişkin belirli istatistiksel indekslere boyun eğmektedir. Bundan dolayı da literatürde, puanlama hatalarına ilişkin kavramsal ve işlevsel tanımları aynı olmayan ve benzer istatistiksel hesaplamaların yapılmasını engelleyen tamamen farklı veri toplama stratejileri kullanan ancak puanlama niteliğinin

ölçütünü aynı düzeyde belirleyen iki veya daha fazla çalışma bulmak oldukça kolaydır (Saal, Downey ve Lahey, 1980).

Oyun sırasında elde edilen istatistikler, özellikle üst düzey yarışmalarda, profesyonel sporlarda, maç boyunca meydana gelen çeşitli sayıda olayın (gol sayısı, penaltı, başarılı şut yüzdesi vb.) sayısını ve sıklığını kayıt altına almayı içermektedir. Bu tür istatistikler, spor performansının çeşitli özelliklerine odaklanmaktadır. Bu istatistiklerin, performansın taktiksel özelliklerini yansıtmayı yansıtmadığını belirlemek çok zor veya imkânsızdır, genellikle bunlar teknik bilgiyi içermektedir. Bu istatistiklerin dayandığı oluşumlar ve olaylar kesin, belirgin hatlara dayanmaktadır. Veri kayıtları ile ilgili gözlemciler arası güvenilirlik iyi bir kanıt olarak görülmemektedir. Oyuncular hakkında karar verilirken genellikle birkaç maçta elde edilen bilgiler bir arada kullanılmaktadır (Grehaigne, 1992).

Bunların yanında, performans değerlendirmeye dayalı bir diğer sınıflandırma ise; kişilerarası karşılaştırmaya, ortak performans ölçütlerine ve standartlarına ve bireysel performans standartlarına dayalı yöntemleri içermektedir (Murphy ve Cleveland, 1991).

Bazı yarışmalarda sporcular, hız, mesafe veya diğer miktarlar gibi fiziksel ölçümler olmadan, sadece uzmanların yer aldığı panellerde puanlar verilerek sıralanmaktadır. Jimnastik, serbest stil kayak ve buz pateni bunlara örnek olarak verilebilir.

Bu durumların tümünde, objektif ölçmeler uygulanamadığından öznel ölçmeler yapmak uygun olan tek yoldur. Buna rağmen, öznel performans değerlendirmelerinde de sorunlar olduğu bilinmektedir. Bu değerlendirmelerin bilinen zayıflıklarından bir tanesi de değerlendiriciler dışında herhangi biri tarafından onaylanamamasıdır. Değerlendirmeyi yapan kişinin ulaştığı kararın altında yatan süreçleri ortaya koymak neredeyse olanaksızdır. Sonuç olarak, öznel ölçümler, kendi amaçlarını takip eden veya kovalayan hakemler veya değerlendirmeyi yapan kişiler tarafından amaçlı olarak değiştirilebilmektedir. Bazen doğru değerlendirme yapmak, bireyin amacı yanında önemsiz kalmaktadır. Öznel değerlendirmeler yanlı olma eğilimindedir. Öznel değerlendirmenin kalitesi, doğruluğu ve yansızlığı, hakemi veya değerlendirme yapan

kişiyi yargılayan sistemin özendiriciliğine bağlıdır. Organizasyon, hakemlerin kararlarını keyfi bir şekilde değiştiremeyecekleri bir mekanizmaya ihtiyaç duymaktadır. Bunun en kolay hatta en popüler yolu, çok sayıda hakem kullanarak, hakemlerin görüşlerini birleştirmektir. Bu bireysel yargıyı engellemekte ve farklı değerlendirmeleri karşılaştırarak “uygun olmayan” değerlendirmeyi belirlemeye olanak sağlamaktadır. Dahası, birden fazla değerlendirmenin birleştirilmesi ile bireylerin heterojen tercihleri ve ölçme hataları düzeltilmektedir. Farklı hakemler, aynı performans için benzer değerlendirmeler yaptığı zaman bu da yakınsak geçerlik (convergent validity) olarak kabul edilmektedir (Murphy, 1982; Murphy ve Cleveland 1991; Saal, Downey, and Lahey 1980).

Öznel performans değerlendirmeleri üzerine ampirik çalışmalar maalesef çok azdır. Spora ilişkin veriler kullanılarak yapılan çalışmalar, spordaki değerlendiricilerin, yani hakemlerin doğru değerlendirme yapmaktansa, hedefleri doğrultusunda değerlendirmeler yapmak için birçok harekete geçirici, özendirici etken olduğunu bulmuştur. Buna rağmen, kendi ülkelerinin sporcularını diğerleri ile karşılaştırmada hakem değerlendirmelerinin objektifliği çok fazla sorgulanmamaktadır. Ancak, hakemlerin sporculara yanlı davranma olasılıkları bulunmaktadır. Bunun çok iyi bilinen örnekleri de bulunmaktadır. 1924 yılında Chamonix’de Çek hakem Çek sporcuyla ilk sıraya yerleştirirken, Avusturyalı hakemlerin ikisi ise Avusturyalı sporcuyla ilk sıraya yerleştirmiştir. Bunun yanında diğer tüm hakemler İsveçli sporcuyla ilk sıraya yerleştirmişlerdir. 1948’de ise St. Moritz’de, tüm hakemler programını başarılı bir şekilde tamamlayan Amerikalı sporcuyla ilk sıraya yerleştirirken, sadece İsviçreli hakem İsviçreli sporcuyla ilk sıraya koymuştur (Wallechinsky, 1991). Buz pateni yarışmalarında sporcuları, hakemlerin kararlarını birleştirerek sıraladığımız zaman, hakemlerin bireysel yanlılıkları son sıralamalarını etkilememektedir (Basset ve Persky, 1994).

Buz pateni, en zarif ve en estetik spordandır. Hızlı hareketleri ve hemen göze çarpmayan, ince değişimleri içermektedir. Buz pateninin bu özellikleri, dalış ve jimnastikte olduğu gibi uygun sıralamanın veya puanlamanın yapılmasını zorlaştırmaktadır. Sürpriz olmamakla birlikte bu üç spor dalında da, sporcuların

sıralarını belirlemek amacı ile uzman hakemler kullanılmaktadır. Bu sorumluluğu tek bir bireye vermek yerine, hakemler paneli kullanılmaktadır. Ancak hakemler paneli kullanılmasına rağmen, hakem puanlarını birleştirmede en iyi birleşimin hangisi olduğu halen zor ve ilginç bir konu olarak karşımıza çıkmaktadır. Bu soruyu dikkate alırsak, hakemlerin puanlarının neden farklı olduğunu sormak mantıklı olabilir. Neden tüm hakemler aynı fikirde değildir?

Buz pateninde panel hakemleri ve başhakem, ulusal federasyonlar tarafından önerilen hakemlerin oluşturduğu uluslar arası hakemler havuzundan her bir yarışma için ISU (International Skating Union/ Uluslararası Buz Pateni Birliği) tarafından seçilmektedir. Her bir performanstan yaklaşık 30 saniye sonra hakemler iki tür puan vermektedirler. Bunlar; teknik ile artistik puanlardır ve bu iki puan türü toplam puanı oluşturmaktadırlar. Her bir puan daha sonra herkesin görebilmesi amacı ile puan tahtasına yansıtılmaktadır. Hakemler ISU tarafından gözlenmekte ve değerlendirilmektedirler. Her bir yarışmayı takiben, hakemler gözlemciler tarafından değerlendirilmektedir, eğer hakemlerden biri paneldeki diğer hakemlerle ciddi bir fikir ayrılığı yaşıyorsa, hakem verdiği bu puanları savunmak zorundadır (Yamaguchi, Ness ve Meacham 1997). ISU gözlemcileri, yarışma sonrası “olay değerlendirme toplantısında” raporu sunmaktadır. Bu rapor hakemlerin hatalarını ve bu hataların kabul edilip edilmediğini içermektedir. Rapor ayrıca, hakemlerin, sporcuların ve antrenörlerin şikâyetlerini de içermektedir. Toplantıda, hakemlerin tümü gözlemcilerin (başhakem), sporcuların, antrenörlerin ve diğer hakemlerin tüm sorularını cevaplamak zorundadırlar. Böylelikle, her bir performans için kabul edilebilir puan aralığı belirlenmekte ve bu aralığın dışına düşen puanlar veren hakemler bunlara kabuledilebilir, makul cevaplar sağlamak zorundadırlar. Toplantıya katılmayanlar veya sorulara cevap veremeyenler cezalandırılmaktadırlar. Eğer bu hakemler para almadan çalışan gönüllülere, uyarı cezası yazılmakta veya bir sonraki yarışmaya katılmaları engellenmektedir (ISU Communication, 1995). ISU, bu hakemleri Olimpiyatlar ve Dünya Şampiyonaları gibi büyük yarışmalara atamayarak da cezalandırmaktadır. Yetersiz olarak değerlendirilen üç tür puan bulunmaktadır (ISU Communication, 1999). Bunlar; sistematik olarak normalden sapan puanlar (ör; belirli ülkelerdeki sporcuların aldıkları yüksek puanlar),

diğer hakemlerin puanlarından oldukça farklı sapmalar ve diğer hakemlerin puanlarından tekrarlı sapmalar.

Sporda öznel değerlendirmeye dayanan karar verme süreçlerine ilişkin çalışmalar incelendiğinde; Garicano, Palacios ve Prendergast (2001, 2005) ile Lidbom ve Priks (2010) futbol hakemlerinin kalabalıktan etkilenecek ev sahibi takımlara avantaj sağlayacak kararlar (ör; maçın sonunda ev sahibi takım 1 gol gerideyse, ortalama süreden daha fazla ek süre eklemek) verdiğini ve maçın önemi arttıkça yanlılığın da arttığını ortaya koymuşlardır. Garicano, Palacios ve Prendergast (2001), futbolda yaptıkları araştırma sonucunda, hakemlerin ev sahibi takım taraftarlarının yani çoğunluğun baskısı altında kaldıklarını ve bunun sonucunda da, oyuncuların sakatlanma sürelerini ev sahibi takımın zaferini destekleyici yönde değiştirdiklerini bulmuşlardır.

Uluslararası Buz Pateni yarışmalarında ise, vatansever kayırmacılığın ortaya çıktığı bulunmuştur. Hakemler, kendi ülkelerinden gelen yarışmacıları kayırmaktadırlar (Campbell ve Galbraith 1996; Zitzewitz 2002).

Bir uçta, her bir buz patencinin performans kalitesini, ölçme hatalarının bir birimi olarak görebiliriz. Bir hakem önemli bir hareketten etkilenebilir veya diğer bir hakem çiftlerden biri sendelediğinde, diğer bireye konsantre olduğu için bunu görmeyebilir. Bu da hakemlerin mükemmel ölçme araçları olmadığını göstermektedir. Burada hakemlerin puanlarının birleştirilmesi bir çeşit ölçme hatasıdır. Bazen, hakemler arasındaki bu farklılıklar ölçme hatalarını temsil etmeyebilir ancak bunun yerine farklı görüşleri temsil edebilir. Hakemlerin tercihleri, gerçek estetik farklılıkları temsil edebilir. Hakemler ulusal onurdan veya diğer daha az ilgili olan etmenlerden etkilenebilirler. Estetiğin karmaşık dünyasında, toplu kararlar vermede sorunlarla karşılaşabiliriz. Prestijli yarışmalarda (Olimpiyatlar veya Dünya Şampiyonaları) hakemler farklı zevklerden ve diğer ölçme problemlerinden daha az etkilenmektedirler. Ancak ulusal veya bölgesel yarışmalarda bu tür problemler daha baskın olmaktadır. Lee (2004) öznel değerlendirmelerin, değerlendirme sistemi tarafından sağlanan teşviklere karşı oldukça hassas olduğunu göstermiştir.

Kullanılan düzenlemeler (puanların hesaplanması, birleştirilmesi), ölçme hataları ve “puanların şişirilmesi” ile etkili bir şekilde başa çıkıldığını göstermesine rağmen amaçlı değişikliklerin veya hilenin yapılmadığı ve hakemlerin sporcuları tesadüfî hatalarla puanladıkları yarışmalarda da bu sistemin iyi çalıştığı bulunmuştur.

Performans ölçümlerinde güvenilirlik ve geçerliği üst sınırlarda kestirebilmek için gerekli olan birkaç tane koşul bulunmaktadır. İlk olarak, hakemlerin hem puanlayacakları performans alanına hem de gerçek puanlama görevine veya sürecine alışık olmaları, bu konuda oldukça deneyimli olmaları gerekmektedir. Buna ek olarak, tüm puanlayıcılar puanlamalarını puanladıkları kişinin aynı hareketlerine, davranışlarına dayandırmalı ve davranışı izler izlemez puanlamayı anında yapmalıdırlar. Son olarak, puanlayıcılar yansız puanlama yapma motivasyonuna sahip olmalıdırlar. Tüm bu ölçütleri karşılayan tek bir senaryo dünya klansmanındaki spor yarışmalarında, sporcuların performanslarını puanlayan hakemlerdir. Dünya Buz Pateni Şampiyonasından elde edilen veriler, öznel değerlendirmelerle ilgili kuramları test etmek için önemli ve eşsiz bir olanak sağlamaktadır. Hakemler farklı performansları gözlemleyip puanlamalar yapmaktadırlar. Buz pateninde, paneli oluşturan hakemler her bir sporcuu iki boyutta puanlamaktadırlar. Uluslar arası büyük organizasyonlarda yıllarca eğitim alan, deneyimli hakemler görev almaktadır. Bu hakemler hem performans alanında hem de puanlama görevinde uzmandırlar. Bunun yanında her hakem tam olarak aynı performansı izlemekte ve puanlamalar standart boyutlara göre yapılmaktadır. Ayrıca her sporcunun performansından hemen sonra puanlar ilan edilmektedir. Böylece, bu durumlar değerlendirmelerinin geçerlik ve güvenilirliğini belirlemek için gerekli olan fırsatları sağlamaktadırlar (Weekly ve Gier, 1989).

1.2. PERFORMANS DEĞERLENDİRMENİN NİTELİĞİ

1.2.1. Geçerlik ve Güvenirlik

Bireyin gerçek yeteneği hakkında çıkarımlar yapmak için değerlendirmeler kullanılmaktadır. Geçerlik ve güvenilirlik, değerlendirmenin en önemli, olmazsa olmaz

öğeleridir (American Educational Research Association, American Psychological Association ve National Council of Measurement in Education [AERA, APA, NCME], 1999; Brown, 1991; Linn ve Burton, 1994). Ancak tüm değerlendirmeler bir miktar ölçme hatası içermektedir. Ölçme hataları pek çok kaynaktan gelebilir. Bunlar: puanlayıcılar, durumlar ve görevler olabilir.

Performans değerlendirme yöntemlerinde, geçerliği ve güvenilirliği sağlamak zor olabilmektedir. Puanlama süreci zaman almakta, farklı maddelerin değerlendirilmesini gerektirmekte ve bu sürece farklı puanlayıcılar katılmaktadır. Bireyin performansı ile ilgili puanlayıcıların görüşlerinde farklılıklar olabilmektedir. Bireyin ortaya koyduğu performansı değerlendirmede bu zayıflıkların üstesinden gelmek gerekmektedir. Gronlund, performans değerlendirme sürecinin geçerlik ve güvenilirliğinin performans değerlendirmenin planlanması sürecinde ele alınması gerektiğini belirtmiştir (Turner, 2003).

Güvenirlik, puanların doğruluğundan ve tutarlılığından bahsetmektedir (Anastasi, 1988; Feldt ve Brennan, 1989; Haertel, 2006). Gözlenen puanın, gerçek puana yakınlığının bir kestirimidir. Klasik test kuramı (KTK) gözlenen puanlarda iki tür varyans kaynağı tanımlamaktadır: (a) gerçek puan varyansı ve (b) hata varyansı. Gerçek puan varyansı, bireylerin yetenekleri arasındaki farkın bir ölçümüdür, hata varyansı ise sistematik değil, tesadüfidir. Güvenirlik, gerçek puanların varyansının gözlenen puanlar varyansına oranı olarak da ifade edilebilir. İyi bir test geliştirmek, hata varyansının kontrolünü gerektirmektedir. Güvenirlik ayrıca bireylerin yeteneklerinin tekrarlanabilirliği ve bunun doğru ve uygun bir şekilde belirlenmesi anlamına gelmektedir. Geçerlik ise test puanlarından yapılan çıkarımların anlamlılığı ve kullanımı ile ilgilidir (AERA, APA ve NCME, 1999; Anastasi, 1988; Kane, 2006). Geçerlik kanıtının üç başlıca türü; (1) ölçüt, (2) kapsam ve (3) yapı geçerliğidir (AERA, APA ve NCME, 1999). Ölçüte dayalı geçerlik veya yordama geçerliği, bir testin bir bireyin belli bir etkinlikteki performansını yordamadaki etkililiğini gösterir. Bu amaç için, bir testte gösterilen performans bir ölçütle kontrol edilir, yani, doğrudan ve bağımsız bir ölçümün yordayıcısı olarak test geliştirilir. Kapsam geçerliği, test maddeleri örnekleminin temsil gücü ile ilgilidir. Bireyin ölçülmek istendiği alanı veya evreni, testin ne kadar iyi temsil ettiğinin bir

göstergesidir. Yapı geçerliği, ölçmek istediğimiz özellik veya yapı ile ilgilidir. Testin ölçmeyi amaçladığı yapıyı ne kadar iyi ölçtüğünü göstermektedir. Yapı geçerliği, karar vermenin temelini oluşturan, test puanlarını ve gözlenen puanların anlamlılığını etkileyen etmenlerin veya faktörlerin belirlenmesini gerektirmektedir. Test geçerliği, özellikle de yapı geçerliği, araştırma ve incelemeleri gerektirmektedir. Yapı geçerliğini belirlemek için üç tane basamaktan söz edilmektedir: (1) ölçülecek değişkenin tanımlanması, (2) ölçülecek olan örtük değişkenin çok veya az olmasının ne anlama geldiğine ilişkin kanıt ve (3) ampirik olarak test edilen denencenin belirlenmesi (AERA, APA, ve NCME, 1999).

Buz pateni, jimnastik, dalış gibi spor dallarının değerlendirmelerinde puanlama ölçekleri oldukça önemli ve popülerdir. Puanlama (dereceleme) ölçeği, hakemin kararlarını veya yargılarını kaydetmek için kullandığı bir ölçme aracıdır. Performansa dayalı değerlendirmede puanlayıcılar/hakemler, puanlama ölçeklerini kullanarak performans değerlendirmelerini nasıl yapacakları konusunda eğitilen uzman olmayan kişiler veya uzmanlardır (Williamson, Mislevy ve Bejar, 2006). Puanlama süreci, içerdiği davranışlara veya boyutlara bağlı olarak puanlayıcılar için zor bir görev olabilir.

Cooper (1981)'a göre puanlama süreci, analiz birimi olarak hizmet eden davranışlar veya hareketlerin gözlenmesi ile başlamaktadır. Gözlenen hareketler, kısa süreli bellekte kodlanmakta ve bir araya getirilmektedir (düzenleme ve birleştirme). Bu aşamada bunlar hakemin kendine özgü özelliklerinden ve dışsal (sosyal veya durumsal) faktörlerden etkilenmektedirler. Zaman içerisinde bu bilgide bozulma, azalma meydana gelecektir. Kodlanan gözlemler, uzun süreli belleğe transfer edilmekte ve orada bir araya gelmekte, toplanmaktadır. Gözlemlerin, bu toplanma sırasında nasıl ağırlıklandırıldıkları net değildir. Fakat genel izlenimden etkilenen yetersiz birleşimler bozulmaya yol açmaktadır. Uzun süreli bellekte de, bu bilgide azalma ve bozulma (tesadüfî veya sistematik) meydana gelmektedir. Puanlama ölçeği verildiği zaman, hakemler uzun süreli bellekten gözlemleri ve izlenimleri geri getirmektedirler. Geri getirmenin niteliği birçok faktöre bağlıdır, örneğin; geçen süre gibi. Daha sonra, hakemler algılanan standartlara göre puanlama kategorileri ile gözlemlerini eşleştirmekte ve puanlar vermektedirler. Bu tür bir puanlama süreci öznel ve gözleme

dayalı yöntemler insanların algılarındaki yanılgılardan etkilenebilmektedir. Böylelikle, puanlama ölçeklerinden elde edilen veriler hatalara ve çarpıtılmaya eğilimlidirler.

Puanlama sürecinin birçok aşamasında hatalar meydana gelebilir. Tesadüfi hata puanlamalarda yanlışlığa ve kusurlara neden olurken, sistematik hata ise puanlamaları belirli bir yönde etkileyebilir. Bu aynı zamanda yanlılık olarak da isimlendirilebilir. Popham (1990) herhangi bir puanlama işleminde üç tane potansiyel hata kaynağı belirlemiştir: ölçme aracı, puanlama işlemi ve puanlayıcılar. Eğer ölçme aracı anlamsız ifadelerden oluşuyorsa, ölçekteki puanlama kategorileri puanlayıcı için yeteri kadar açık değilse veya puanlayıcı puanlayacağı yapıyı anlamıyorsa, bu ölçme aracı uygun olmayan puanlamalara yol açacaktır. Puanlama işlemindeki değişiklikler (çevrimiçi puanlama sistemine geçiş), günlerce süren puanlama işlemleri ve puanlayıcının aynı anda birden çok özelliği değerlendirmesini gerektiren işlemler, performansı puanlamanın doğruluğunu etkilemektedir. Son olarak, puanlayıcılar potansiyel hata kaynağı olarak davranabilirler. Test maddelerinden farklı olarak, hakemler zaman içerisinde ve puanladıkları bireylere göre farklılaşabilirler. Yapı ile ilgili olmayan özellikler bireyin değerlendirmesini etkileyebilecek potansiyele sahiptir. Scullen, Mount ve Goff (2000) puanlayıcı etkisini “performansı puanlamada sistematik varyans ile sonuçlanan etkiler kategorisi” olarak tanımlamıştır. Böylece, puanlayıcılar bir değişkenlik kaynağı olarak ele alınmalıdır.

Puanlayıcı etkisi, literatürde (eğitimsel ölçmeler, klinik değerlendirmeler, performans değerlendirmeleri ve seçim yapma) geniş olarak yer bulmaktadır. Puanlayıcılar bir bilgi işlemci gibi hareket etmekte ve gözlem, saklanmış bilgiyi geri getirme, bunu organize etme, birleştirme, ağırlıklandırma ve bütünleştirme sonucu bir puan vermeyi gerektiren karmaşık ve çok yönlü bir süreçten geçmektedirler (Myford ve Wolfe, 2003). Bu süreç karmaşık olduğu kadar, hata eğilimlidir (Cronbach, 1990). Test puanlarındaki değişkenlik uzun zamandır hakemlerle ilişkilendirilmektedir (Cason ve Cason, 1984). Puanlayıcılar değerlendirme sürecinde kullanıldığı zaman, değişkenlik kaynağı olmakta, ölçme hatalarına yol açabilmekte ve değerlendirmenin güvenilirliğini düşürebilmektedirler. Hata, tesadüfi veya sistematik olabilir. Tesadüfi hatanın

belirlenmesi zor olduğundan dolayı bu düşük güvenilirlik (hakemin güvenilmezliği) ile sonuçlanabilir.

Scullen, Mount ve Goff (2000)'a göre; sistematik hata, puanlama desenine bağlı olabilir. Tüm hakemlerin, tüm bireyleri değerlendirdiği bir desende, beklenen sistematik etkiler bireyler arasında dengelenmekte veya temizlenmektedir. Sonuç olarak, bu tür çaprazlanmış desenlerde güvenilirlik katsayısı hesaplanırken bu etkiler dikkate alınmaktadır (Raymond ve Viswesvaran, 1993). Buna rağmen, büyük ölçekli performans değerlendirmelerinde yuvalanmış puanlama desenleri kullanılmaktadır çünkü daha az maliyetle ölçme işlemi yapılabilmektedir. Yuvalanmış desenlerde, tüm hakemler tüm bireyleri puanlamak yerine, hakemlerden oluşan alt gruplar belirli bireyleri puanlamaktadırlar (Raymond ve Viswesvaran, 1993). Böylelikle, puanlayıcı etkisi, bireyler arasında farklılaşmaktadır. Böyle durumlarda, hakemler tarafından getirilen sistematik varyansın dikkate alınmasında yarar vardır. Araştırmalar, hakemlerin öznel hatalarının ciddi bir biçimde puanlamadaki varyansa katkıda bulunduğunu ortaya koymuşlardır (Conway, 1996; Viswesvaran, Ones ve Schmidt, 1996). Performans değerlendirmede hakemlerin kullanımı oldukça yaygındır. Bu nedenle, tüm puanlayıcı etkilerinin incelenmesi ve hataların test puanları ile yorumları üzerindeki etkilerinin belirlenmesi oldukça önemlidir.

1.3. PUANLAYICI HATALARININ TÜRLERİ

Çok fazla sayıda puanlayıcı hatası bulunmasına rağmen, bunlardan en çok bilinen dört “klasik” puanlayıcı hatası Saal, Downey ve Lahey (1980) tarafından sınıflandırılmıştır. Bunlar; (1) cömertlik/katılık, (2) merkezi eğilim, (3) ranj daralması (ranjin kısıtlanması) ve (4) halo. Bu bölümde bu dört puanlayıcı hatasının yanı sıra diğer daha az bilinen hatalar da ele alınacaktır. Bunlar ise; (5) doğruluk/yanlışlık (accuracy/inaccuracy), (6) tutarsızlık, (7) mantıksal hata, (8) yakınlık (proximity), (9) yenilik, yakın zamanda meydana gelme (recency), (10) öncelik (primacy), (11) algı farklılıkları, (12) sıra etkisi, (13) taşıma, devam etme (carry over) ve (14) zıtlık (contrast) (Iramanareet ve Yudkowsky, 2007; Myford ve Wolfe, 2003). Bu hata türlerine ek olarak, farklı

puanlayıcı hataları birbirine geçmiş olabilir. Ancak, bu hataları ölçme ve belirleme yolları da birbiri ile örtüşmektedir (Myford ve Wolfe, 2003).

1.3.1. Cömertlik/Katılık

Puanlayıcıların katılık derecelerindeki değişimler puanlayıcının cömertliği/katılığı olarak bilinmektedir. Bu dereceler puanlayıcıdan puanlayıcıya, zaman içerisinde ve gruptan gruba değişiklik göstermektedir (Lumley ve McNamara, 1995; Lunz, Wright ve Linacre, 1990; Wilson ve Wang, 1995). Bu hata, puanlayıcının yapabileceği en ciddi etkilerden biridir. Cömertlik/katılık, puanlayıcının bireyin gerçek performansına istinaden beklenen puandan daha yüksek (veya daha düşük) puan verme eğilimi olduğu zamanlarda ortaya çıkmaktadır. Kingsbury, yüksek/düşük uyumsuzluklara ilk olarak odaklanan kişidir. Kneeland, cömertlik terimini “kullanılan ölçeğin orta noktasının oldukça üzerinde puanlama yapan” hakemler için kullanmıştır. Kneeland’ın terminolojisine dayanarak, Ford, ölçeğin düşük ucunda puanlama yapanlar için de katılık terimini kullanmıştır. Guilford, cömertlik hatasını tanımlamış ve bu puanlayıcıları “kolay puanlayıcılar” veya “zor puanlayıcılar” olarak tanımlamıştır. Puanlayıcının katı veya cömert olması sabit hatalar altında tanımlanmaktadır (Aiken, 1996). Bireyin davranışından bağımsız olarak puanlayıcının durağan ve sabit puanlama eğilimini yansıtmaktadır.

Cömertlik/katılık en sık çalışılan etkilerden bir tanesidir ve bu etkiyi tanımlamada farklı yöntemler kullanılmaktadır. Bunu belirlemenin en kolay yolu puanlamanın dağılımına bakmaktır. Pozitif çarpık dağılımlar hakemin katılığını temsil ederken, negatif çarpık dağılımlar da hakemin cömertliğini temsil etmektedir. Bunun yanında, bir puanlayıcının ortalama puanını, ölçeğin orta noktası veya diğer puanlayıcıların genel ortalamaları ile karşılaştırarak da bir yargıya varılabilir. Ölçeğin orta noktasından veya diğer puanlayıcıların genel ortalamalarından her iki yöne doğru anlamlı sapmalar, puanlayıcının katılığı veya cömertliğini göstermektedir. Buna alternatif bir yaklaşım da varyans analizi çerçevesinde anlamlı hakem etkisine bakmak olabilir (Myford ve Wolfe, 2003).

Çaprazlanmış desenlerde (tüm bireylerin tüm hakemler tarafından puanlandığı desenler) norm dayanaklı karşılaştırmalar kullanıldığında hakemlerin cömertliği/katılığı bireylerin görelî sırasını deęiřtirmeyecektir. Bu etki tüm bireyler tarafından deneyimleneceğinden durum dengelenecektir. Puanların ortalaması alındığında, beklenen puanlar dağılımında sadece sabit bir kayma olacaktır. Böylelikle, cömertlik/katılık dışında başka bir puanlayıcı hatası karışmazsa, norm dayanaklı karşılaştırmaların kullanıldığı çaprazlanmış desenler güvenilir olacaktır. Bu durum, yuvalanmış desenler (belirli bir hakemler grubunun belirli bir birey grubunu puanladığı desenler) için doğru deęildir. Yuvalanmış desenler ekonomik olduklarından dolayı birçok büyük ölçekli deęerlendirmelerde yaygın olarak kullanılmaktadırlar. Bu tür desenlerde, puanlayıcıların cömertliği/katılığı, puanlama yapan puanlayıcı kümesine baęlı olarak bireyin görelî sırasını etkilemektedir. Bireyler üzerinde puanlayıcıların katılık veya cömertlik etkileri farklı olabilmektedir. Bu etki, puanlama yapan hakem sayısını artırarak en aza indirilebilir. Dięer yandan, ölçüt dayanaklı çerçeveden bakıldığında, puanlayıcı cömertliği/katılığı tam tersi bir etki ortaya koymaktadır. Her iki desende de hakemin cömertliği/katılığı kesme noktası altında ve üzerinde olan bireylerin oranlarını deęiřtirmektedir (Wolfe, 2004).

1.3.2. Ranj Kısıtlanması (Ranj Daralması)

Bu olgu puanlayıcıların ölçeęi kullanmada başarısız oldukları durumları temsil etmesinin yanında, tüm puanları ölçeęin belirli bir bölümüne veya kısmına sınırlaması durumunu da temsil etmektedir. Ölçeęin bir bölgesine yığılan puanlar deęişkenliği sınırlamakta ve puanlamayı daha az ayırt edici kılmaktadır (Wilson ve Case, 2000). Bu küme, ölçeęin uçlarında oluşabildięi (aşırılık) gibi orta noktalarında da oluşabilir ve bu da merkezilik olarak adlandırılmaktadır (Wolfe, 2004). Ranj daralması literatürde, puanlayıcıların sabit tepkilerinin bir sonucu, düşük farklılaşma/ayırt etme eğilimi olarak da ele alınabilmektedir.

Puanlar ölçeęin orta noktası etrafında kümeleniyorsa bu merkezi eğilim olarak bilinmektedir. Merkezi eğilim, “tüm puanlama nesnelere orta noktada veya puanlama doğrusunun merkezinde puanlama ve uçlardaki puanları kullanmama eğilimi” olarak

tanımlanabilir. Ortalama puanlama, ölçeğin orta noktası ile uyum gösterdiği zaman, bu ölçeğin orta noktası etrafında puanların kısıtlanmasını göstermektedir (Saal, Downy ve Lahey, 1980). Ortalama puanlar, ölçeğin orta noktasından daha yüksek (cömertlik) veya daha düşük (katılık) olabilir, bu da ölçeğin diğer kısımlarında puan aralığını (ranjını) kısıtlamaktadır. Böylece, merkezi eğilim ranj kısıtlamasının özel bir halidir. Cömertlik ve katılık, puanların ölçeğin iki ucuna yığılması (düşük/yüksek), ranj kısıtlamasının özel bir durumu olarak görülebilir (Myford ve Wolfe, 2003; Wilson ve Case, 2000)

Puanlayıcıların uçlarda puanlama yapmaktan kaçınmalarının çeşitli nedenleri olabilir. Örneğin; puanlayıcılar katı veya cömert olma ile ilişkili olan çekişmeden ve anlaşmazlıklardan kaçınmak için tarafsız veya etkisiz kalmayı tercih etmekteydiler. Ayrıca, eğer puanlayıcılar puan verdikleri birey ile kişisel olarak iletişime geçeceklerse katı değerlendirmenin sebep olacağı kötü görünümünden kaçınmak istemektedir. Puanlayıcıların puanlama ölçeklerini nasıl kullanacaklarını anlamamaları da olasıdır ve sonuç olarak puanlamalar şüpheli olmaktadır. Puanlayıcıların bir bölümünün farklı performans düzeyleri arasında ayırım yapamamaları da bir tür başarısızlıktır (Wilson ve Case, 2000).

Saal, Downy ve Lahey (1980) ranj kısıtlaması etkisini çalışmak için üç tane yaklaşım belirlemişlerdir. Bunlardan en basit olanı, bir puanlayıcının belirli bir özelliğe tüm bireylere verdiği puanların standart sapmasını belirlemektir. Küçük standart sapmalar, daraltılmış varyansı ve ranjin iyice daraldığını göstermektedir. Ortalama puanları ve bu puanların ölçeğin orta noktasına yakınlığını gözlemleyerek, ranj kısıtlaması merkezi eğilimin bir temsili olarak belirlenebilir. Buna alternatif bir yaklaşım da puanlamaların sıklık (frekans) dağılımındaki tepe noktalarını gözlemlemek olabilir. ANOVA çerçevesinden bakıldığında, puanlayıcıx bireyx özellik etkileşimindeki anlamlı olmayan birey etkisi, bireyler ve performans düzeyleri arasında farklılığın olmadığını göstermektedir.

Diğer puanlayıcı hataları ile benzer olarak, bu etkinin üstesinden gelebilmek için puanlayıcı eğitiminde puanlama ölçeklerinin kullanımının üzerine gitmek, farklı performans düzeylerini temsil eden örnekler göstermek ve bu tür hatalar yapıldığında

puanlamanın hatalı sonuçlarını puanlayıcılara açıklamak gerekmektedir (Ramineni, 2008).

1.3.3. Halo Etkisi

Halo en çok bilinen ve çalışılan puanlayıcı etkilerinden bir tanesidir (Iramanareet ve Yudkowsky, 2007; Myford ve Wolfe, 2003). Halo, bireyin genel izleniminin puanlayıcının puanlama davranışını etkilemesi anlamına gelmektedir. Örneğin, puanlayıcı bireyin görünüşünden etkilenebilir.

Fisicaro ve Lance (1990) etki kaynağına dayanarak halo etkisini üç grupta toplamışlardır: (1) puanlayıcının birey davranışlarını ayırt edememesi, (2) genel izlenimin etkisi ve (3) bireyin performansının dikkat çeken, belirgin özelliklerinin etkileri. Puanlayıcının yargılarını bozan ilgisiz ölçütleri de içermektedir. Bu etki, ilk kategoriye göre puanlayıcının bireyin davranışını potansiyel bağımsız boyutlara ayıramamadaki başarısızlığı olarak tanımlanmaktadır. İkinci kategoriye göre, halo etkisi puanlayıcının bireyin performansını tek ve potansiyel olarak bağımsız ya da yarı-bağımsız boyutlarda puanlaması sırasında, bireyin genel izleniminden etkilenmesi olarak tanımlanabilir. Anastasi (1988) ise bu puanlama etkisini bir özelliğin diğerleri üzerindeki etkisi olarak tanımlamıştır. Yani üçüncü kategori olan çarpıcı boyut modeline denk gelmektedir.

Halo etkisi bireyin farklı özelliklerine ilişkin puanları arasındaki korelasyonların hesaplanması ile belirlenebilir. Farklı olduğu bilinen (benzemeyen) bir boyutta eğer korelasyon yüksek ise bu yüksek korelasyon halo etkisi olarak yorumlanabilir (Keaveny ve McGann, 1975). Ancak yüksek korelasyonları halo etkisinin bir kanıtı olarak yorumlamada dikkatli olunmalıdır. Halo olumlu ve olumsuz olarak tanımlanmaktadır. Örneğin; puanlayıcının belirli bir bireye karşı “çok kolay” veya “çok zor” olma eğilimi. Olumlu halo ilk değerlendirmelerdeki olağanüstü puanların sonucunda tüm özelliklerin avantajlı puanlanmasıdır. Olumsuz halo ise ilk değerlendirmelerdeki düşük puanların sonucu olarak çoğunluğun düşük değerlendirilmesidir (Aiken, 1996; King, Hunter ve Schmidt, 1980).

Saal, Downy ve Lahey (1980) korelasyon, faktör analizi, ANOVA ve dağılım gibi farklı yaklaşımlara dayanarak halo etkisinin işlemsel birçok tanımını ileri sürmüşlerdir. Farklı özelliklerin puanlanmasındaki yüksek korelasyonlar halo etkisini yansıtmaktadır. Özellikler için, temel bileşenler analizi veya faktör analizinden elde edilen birkaç (daha az) faktör de halo etkisinin kanıtı olabilir. Puanlayıcı-birey birleşimleri (bir diğer deyişle azaltılmış bireye ait varyans) için sınırlandırılmış dağılım halo etkisini belirlemede kullanılan bir diğer yaklaşımdır. ANOVA çerçevesinden bakıldığında ise, varyansın önemli bir miktarını açıklayan anlamlı puanlayıcı-birey etkileşimi de halo etkisini göstermektedir.

Halo, yeterli sayıda ve temsil edici miktarda davranış örnekleme seçilerek; puanlayıcılar için açıkça farklılaştırılan, ayırt edilebilen özellikler kullanılarak; belirli bir zaman içerisinde bir puanlayıcıya tüm bireyleri tek bir özellik açısından puanlatarak; puanlanacak özellik sayısını azaltarak ve puanlayıcıları eğitim sırasında halo etkisi ve onun etkileri hakkında bilgilendirerek giderilebilir (Myford ve Wolfe, 2003).

1.3.4. Doğruluk/Yanlışlık (Accuracy/Inaccuracy)

Puanlayıcı doğruluğu, gözlenen puanlar ve gerçek puanlar arasındaki tutarlılık olarak tanımlanmaktadır. Gerçek ve gözlenen puanlar arasındaki artıklar doğruluğun bir ölçümüdür (Wolfe, 2004). Gerçek puanlar, yetenek parametresinin bir kestirimi olan beklenen puanlardır veya uzman puanlayıcılar grubundan elde edilen standart ölçüt puanlardır (Myford ve Wolfe, 2003). Doğruluk ve puanlayıcı hataları ters olarak ilişkilidirler ve tüm puanlayıcı hatalarına ilişkin indeksler puanlayıcı doğruluğunun doğrudan olmayan belirleyicisidirler (Murphy ve Balzer, 1989).

1.3.5. Tutarsızlık

Dört “klasik” puanlayıcı hatası (cömertlik/katılık, halo, merkezi eğilim ve ranjin kısıtlanması) biribine benzemekte ve puanlamalarda sistematik hatayı göstermektedir. Ancak tek biçimli olmayan hatalarda bulunmaktadır. Bu hatalar zaman içerisinde,

özellikler arasında ve/ya gruplar arasında, alt gruplarda ve bireyden bireye değişmektedir. Bu hatalar puanlayıcıların değerlendirmeleri arasında rastgele değişkenlik şeklinde görülmektedir. Bu hata puanlayıcıların tutarsızlığı veya rastgelelik etkisi olarak bilinmektedir (Iramanareet ve Yudkowsky, 2007; Myford ve Wolfe, 2002). Tutarsızlık, puanlayıcıların ölçeği veya ölçek kategorilerini tam olarak anlayamamalarının bir sonucu olabilir (Downing ve Haladyna, 2004). Ayrıca puanlama sürecinde kendiliğinden ortaya çıkan dinamikler sonucunda da olabilir, örneğin; puanlayıcı-birey etkileşimi. Bu rastgelelik farklı yetenek düzeyinde olan bireylerin beklenmedik bir şekilde normal olmayan puanlar alması ile sonuçlanabilir. Yüksek yetenek düzeyine sahip bireyler düşük puanlar alabilir veya tam tersi söz konusu olabilir (Iramanareet ve Yudkowsky, 2007).

1.3.6. Mantıksal Hata

Halo etkisine benzer bir etki de mantıksal hata olarak bilinmektedir. Puanlayıcılar, özelliklerin benzerliklerine ilişkin önyargılarına dayanarak farklı özelliklere benzer (yüksek korelasyon gösteren) puanlar vermektedirler (Aiken, 1996; Myford ve Wolfe, 2003). Bu durumlarda özelliklerin kavramsal benzerliği gerçek olmamakla birlikte, bu benzerlik puanlayıcının kişisel inancının bir sonucudur. Diğer bir deyişle, “puanlayıcının zihninde olan mantıksal varsayımlara” yüklenmektedir. Halo etkisi ve mantıksal hata arasındaki fark; halo etkisi puanlayıcıların bireyin farklı özellikleri arasında ayırım yapma yetersizliği sonucu ortaya çıkarken, mantıksal hata ise genel olarak, bireyi hesaba katmadan özellikleri ayırt edememe yetersizliğini yansıtmaktadır (Myford ve Wolfe, 2003).

1.3.7. Yakınlık (Proximity Error)

Yakınlık hatası, puanlama formunda özellikler arası fiziksel mesafenin bir fonksiyonudur. Yakınlık hatası, farklı özelliklerin puanlanması arasındaki yakınlığın etkisini göstermektedir. Bu olgu gerçek olmayan bir şekilde özellikler arası yüksek korelasyonları göstermektedir. Bu hatanın bir diğer tanımı da “iki özelliğin

puanlanmasında uzayda veya zamanda yakınlık” ve puanlama formunda birbirine yakın olan veya birbiri ardına gelen maddeleri benzer şekilde puanlama eğilimidir (Aiken, 1996).

1.3.8. Yenilik, Yakın Zamanda Meydana Gelme/ Öncelik (Recency / Primacy Error)

Yakın zamanda meydana gelme etkisini, puanlayıcının daha sonra meydana gelen performansı, daha önce meydana gelene (zaman içerisinde) göre daha ağırlıklı olarak puanlama eğilimi şeklinde tanımlamaktadır. Bireyin bıraktığı en son etki toplam ölçüm periyodunu gölgelemektedir (Myford ve Wolfe, 2003). Aiken (1996) bu hatayı en yakın zamanda meydana gelen performans hatası olarak ifade etmektedir. Bireyin aldığı puan, tüm davranışlarına karşın en çok en yakın performansından etkilenmektedir. Bireyin en son meydana gelen performansı bazı sebeplerden dolayı normalden oldukça farklıysa, böyle bir puanlama daha fazla hata barındıracaktır. Diğer yandan, öncelik etkisi, puanlayıcının ilk etkiye göre puan verme eğilimidir. Bu gibi durumlarda puanlayıcının ilk performansa ilişkin görüşü tüm puanlama sürecine hâkim olmaktadır.

1.3.9. Sıra Etkisi (Order Effects)

Bireyin puanlanma sırasına bağlı olarak ortaya çıkmaktadır. Kompozisyonları puanlama sırasında, ilk puanlanan kompozisyonların yüksek puanlar aldığı, daha sonrakilerin ise daha düşük puanlar aldığı bulunmuştur (Coffman ve Kurfman, 1968). Bu, puanlayıcının yorgunluğu veya puanlayıcının standartlarındaki zaman içerisindeki değişimler sonucu ortaya çıkmış olabilir. Ayrıca, puanlayıcının puanlama ortamı ve puanlama işlemine daha aşina olması sonucunda oluşabilir (Ramineni, 2008).

1.3.10. Algı Farklılıkları (Perception Differences)

Puanlayıcıların puanlama senaryosuna getirdiği geçmiş deneyimlerine dayanan bir bakış açısı olarak düşünülebilir. Algı farklılıkları, puanlayıcıların inançları, tutumları, kişilik

özellikleri ve demografik özelliklerine bağlı olarak gelişmektedir (Myford ve Wolfe, 2003).

1.3.11. Taşıma, Devam Etme (Carryover Effects)

Çaprazlanmış desenlerde, katılımcıların yer aldığı farklı koşullarda etkilerin devam etmesi beklenmektedir. Bunlar, hafıza, uygulama veya yorgunluk olabilir. Taşıma etkisi zarar verici olabilmekte, desenin içsel ve dışsal geçerliği için bir tehdit oluşturmaktadır (Aiken, 1996). İki tane etki belirlenmiştir. Bunlardan bir tanesi aynı testte maddeden maddeye taşıma etkisi, bir diğeri de bir testten diğerine taşıma etkisidir. İlk durumda, aynı öğrenciyi değerlendirirken puanlayıcı ilk sorudan diğerine bu etkiyi taşıyabilir. İkinci durumda ise, puanlayıcı bir kişiden diğerine bu etkiyi taşıyabilir. Myford ve Wolfe (2003), yaptıkları geniş kapsamlı araştırmalar sonucunda bu etkinin yakınlık etkisi ile benzer ve ilişkili olduğunu bulmuşlardır.

1.3.12. Zıtlık (Contrast Error)

Murray, zıtlık etkisini, puanlayıcının puanlama yaptığı özellikte kendisini, değerlendireceği birey ile karşılaştırma ve bireyi ters yönde puanlama eğilimi olarak tanımlamıştır. Örneğin, puanlayıcı çok düzenliyse, bireyi kendi ile karşılaştırdığı zaman daha düzensiz bulacak ve buna bağlı olarak daha düşük puan verecektir. Guilford, zıtlık etkisinin aynı zamanda benzerlik etkisi olarak da ortaya çıkabileceğini ve yanlılığa yol açabileceğini belirtmiştir. Örneğin, düzensiz bir puanlayıcı, düzensiz bir bireyi puanlarken, kendi ile benzerlikten dolayı bunu farketmeyebilir ve buna bağlı olarak doğru olmayan bir şekilde puanlama yapabilir. Zıtlık etkisi, puanlayıcının beklenenden daha yüksek (düşük) puanlama yapma eğilimidir (Aiken, 1996).

1.4 ÖLÇME HATALARINI BELİRLEMEDE KULLANILAN YAKLAŞIMLAR

Bireyin yeteneği ile ilgili güvenilir çıkarımlar yapabileceğimiz değerlendirmeler yaratabilmek için bu hata kaynakları azaltılmalıdır. Ölçme hatalarının niceliğini

belirlemek için çerçeveler geliştirilmiştir. Bunlardan yaygın olanları; klasik test kuramı (KTK), genellenebilirlik kuramı (G-Kuramı) ve çok değişkenlik kaynaklı Rasch modelidir (ÇDKRM) (Engelhard, 1996).

1.4.1. Klasik Test Kuramı

Klasik test kuramı, gözlenen puan, gerçek puan ve hata puanı kavramlarına dayanır. Klasik test kuramında gerçek puan, sayılabilir sonsuz çokluktaki ölçme sonuçlarının aritmetik ortalaması olarak ifade edilmektedir. Bu ölçme sonuçlarında bulunan hataların evrendeki dağılımının aritmetik ortalamasının sıfıra eşit olduğu sayılına dayanılarak gözlenen puanın, sonsuz çokluktaki ölçme sonuçlarının ortalaması alındığında gerçek puana eşit olduğu söylenir. Bu anlamda gerçek puan teorik bir kavramdır. Gözlenen puan, gerçek puan ile hata puanı cinsinden, $X=T+E$ eşitliği ile ifade edilir ve bu ifade klasik test kuramının temel denklemi olarak adlandırılır (Lord ve Novick, 1968).

KTK'da güvenilirlik kuramının temel sayılı ölçme hatalarının "tesadüfî" olduğudur. Herhangi bir kişi için, ölçmedeki bir hata tamamen tesadüfî (random) bir olay değildir. Fakat çok sayıda kişiler üzerinde yapılan ölçümlerde, ölçme hatasının sebepleri oldukça çeşitli ve karmaşık olarak varsayıldığından dolayı ölçme hataları tesadüfî değişkenler olarak tanımlanır (Crocker ve Algina, 1986; Murphy ve Davidshofer, 2001).

Eğer hatalar tesadüfî değişkenlerin gerekli niteliklerini taşıyorsa hataların pozitif ve negatif olmada eşit şansa sahip olduğunu ve hataların gerçek değerle veya başka testlerdeki hatalarla ilişkisiz olduğunu varsaymak gerekir. Yani;

1. Ölçme hatalarının ortalaması=0
2. Gerçek puanlarla hatalar arasında ilişki yoktur: $r_{te} = 0$
3. Farklı ölçümlerden gelen hatalar arasında ilişki yoktur: $r_{e1e2} = 0$

Bu üç temel sayılıya dayanılarak, güvenilirlik kuramı geliştirilmiştir (Lord ve Novick, 1968).

Birçok performans değerlendirme ortamında yaygın olan uygulamalardan bir tanesi de bireyin performansının iki puanlayıcı tarafından puanlanmasıdır. Eğer iki puanlayıcı verdikleri puanlarda fikir ayrılığına düşerse yani uyum göstermezse üçüncü puanlayıcı kullanılmaktadır. Böylece, puanlayıcılar arası güvenilirlik, puanların güvenilirliğinin bir kanıtı olarak hesaplanmaktadır. İki tür puanlayıcılar arası güvenilirlik kestirimi bulunmaktadır; uyum kestirimi ve tutarlık kestirimi (Stemler, 2004). Klasik Test Kuramında puanlayıcılar arası güvenilirliğin belirlenmesinde kullanılabilen istatistiksel teknikler; basit yüzde, Cohen'in Kappa Katsayısı, Kendall Uyuşma Katsayısı, Pearson Korelasyon Katsayısı, Sınıf İçi Korelasyon Katsayısı ve ortalamaların karşılaştırılmasıdır.

Uyum kestiriminde, iki puanlayıcının puanları arasındaki tam ve/ya yaklaşık uyumun (± 1) ölçüsü kestirilmektedir. Eğer puanlayıcılar ölçülen yapıyı benzer şekilde yorumlamışlarsa puanlayıcıların aynı veya yakın puanlar ürettiklerine inanılmaktadır. Bu da puanların yorumunun geçerliği için önemlidir. Uyum derecesi basit uyum yüzdesi şeklinde hesaplanmaktadır. Basit yüzde, iki puanlayıcının veya aynı puanlayıcıların verdikleri puanların uyumunun basit yüzdesinin hesaplanmasına dayanmaktadır. Bir puanlayıcının puanı, diğer puanlayıcının puanından 1 puan yüksek veya düşük olmadığı sürece, puanlayıcıların uyumlu oldukları düşünülmektedir. Sınırlı sayıda puanlama kategorisi olduğunda veya seçilebilecek tüm noktalar birbirine yakın olduğunda örneğin; 4'lü puanlama ölçeğinde (1'den 4'e kadar), yaklaşık uyumun (adjacent agreement) kestirimi normalden fazla olabilir. Puanlayıcılar arası uyumun kabul edilebilir sınırı %70 veya daha büyüktür (Brown, Glasswell ve Harland, 2004). Cohen Kappa Katsayısı, şans düzeyini dikkate alarak uyum düzeyini kestirmeyi sağlar. Kappa istatistiği iki veya daha fazla puanlayıcının yaptığı değerlendirmeler arası uyumun belirlenmesi için kullanılır (Goodwin, 2001). Kendall Uyuşma Katsayısı, çok sayıda puanlayıcının performansı değerlendirmede vermiş olduğu puanlar arasında ne ölçüde uyum bulunduğunu belirler. Parametrik olmayan istatistiksel bir tekniktir (Howell, 2002).

Tutarlılık kestirimi, sürekli olan verilerde daha kullanışlıdır. Brown, Glasswell ve Harland (2004)'ın açıkladığı gibi; yüksek katsayılar hakemlerin performansları

puanlamada yüksek ve düşük puanları benzer örüntüler içerisinde kullandığını göstermektedir. Buna rağmen, her bir hakemin ortalama puanı oldukça farklı olabilir. Çünkü bir hakemin verdiği yüksek puan, diğer hakemin verdiği puan ile eşit olmak zorunda değildir. Böylece, hakemler verdikleri puanlar farklı olsa da, eğer bireyleri aynı şekilde sıralıyorlarsa birbirleri ile tutarlı görünmektedirler. Bunu bir örnek üzerinde açıklamak gerekirse, katı bir puanlayıcının 20 kişiyi puanladığını, diğer hakemin ise aynı 20 kişiyi cömert puanlama sonucunda 1'er puan daha yüksek puanladığını düşünelim. Eğer hakemler bireyleri benzer bir örüntü içerisinde aynı şekilde sıraladıysa, tutarlılık katsayısı yüksek olacaktır. Bu yaklaşımın bir diğer zayıf noktası ise, hakemler arasında küçük farklılıklar olsa bile tutarlılık katsayısının düşmesidir. Böyle bir durumda ise, puanlayıcılar arası tutarlılığın zayıf olduğuna ilişkin yanlış bir sonuca ulaşılabilir, fakat bunun nedeni ranjin kısıtlanması (darltılması) olabilir. Uyum kestirimindeki gibi %70 ve daha büyük değerler puanlayıcılar arası tutarlılığı göstermektedir. Puanlayıcılar arası güvenilirliği belirlemek için kullanılacak bir diğer yöntem, en çok kullanılan ve yaygın olarak bilinen korelasyon hesaplamasıdır (Anastasi ve Urbina, 1997). Tutarlılığın belirlenmesinde yaygın olarak kullanılan iki korelasyon katsayısı bulunmaktadır. Veriler aralık ölçeğinde ise, puanlayıcılar arası tutarlılık derecesini belirlemekte Pearson Momentler Korelasyon katsayısı kullanılmaktadır. Pearson Korelasyon Katsayısı iki puanlayıcının verdikleri puan dağılımı arasındaki ilişkiyi belirlemektedir. Ancak ikiden daha fazla puanlayıcının olduğu durumlarda, varyans analizine dayalı olarak hesaplanan grup içi (intra-class) korelasyon katsayısı hesaplanmaktadır (Baykul, 2000; Goodwin, 2001). Eğer veriler sıralama ölçeğindeyse, puanlama ölçeğindeki iki puanın aralığı eşit olmayacağından, Spearman sıra korelasyon katsayısı kullanılmaktadır. Pearson veya Spearman puanlayıcılar arası güvenilirlik katsayısı, puanlayıcılar arası puanlama hatalarının neden olduğu ölçme hatalarının miktarının nicel kestirimini vermektedir (Stemler, 2004).

Puanlayıcılar arası güvenilirliğin hesaplanmasında, puanlayıcıların bireylerin ölçülen özelliklerine verdikleri puanların ortalaması ve toplamının alınması olanaklıdır. Ortalamaların karşılaştırılmasında t-testi veya varyans analizi kullanılabilir (Baykul, 2000).

Performans değerlendirme birçok alanda karşımıza çıkmaktadır. Yüksek öneme sahip durumlarda, bir hakemin oldukça yüksek puanladığı bir bireyi diğer hakemin çok düşük puanlaması çok yaygın görülmemektedir. Çünkü iyi performansın gerekliliklerini anlamada genellikle iyi eğitilmişlerdir ve hakemlerin mantıklı olması beklenmektedir. Bu nedenle %70 uyum değeri yeterli olmayabilir. Bu gibi durumlarda daha yüksek değerler beklenmektedir. Burada önemli olan, araştırmacıların puanlayıcılar arası güvenilirliği değerlendirmede uyum veya tutarlık katsayılarının sınırlılıklarından haberdar olmalarıdır. Puanlayıcının performansını değerlendirmede tek bir ölçüte bağlı kalınmamalıdır (Stemler, 2004).

Hem uyum kestiriminde hem de tutarlık kestiriminde, puanlayıcıların benzer puanlar vermeleri veya aynı sıralama örüntüsüne sahip olmaları arzu edilmektedir. Puanlayıcılar arası anlaşmazlıklar (farklı puanlama örüntüleri) hata olarak düşünülmektedir (Crocker ve Algina, 1986).

KTK'nın sınırlılığı hata varyanslarını homojen olarak ele almasıdır. Diğer bir deyişle, farklı hata kaynakları birbirinden ayıramamakta ve tek seferde sadece bir tane hata kaynağı belirlenebilmektedir. Diğer tüm hata kaynakları, hata kestiriminin bir parçası veya gerçek puan varyansı olarak ele alınmaktadır (Fan ve Chen, 2000). Bu durumda, farklı hata kaynaklarının güvenilirliği ne ölçüde etkilediğini araştırmak imkânsız olmaktadır. Puanlayıcılar arası uyum ve tutarlık kestirimlerinin hesaplanması oldukça kolaydır. KTK'nın sınırlılıklarından dolayı, bu kestirimler sonucunda elde edilen bilgiler yetersizdir ve genellikle performans değerlendirmede puanlar yanıltıcı sonuçlar verebilir. Puanlayıcılar birbirileri ile anlaşmaya zorlandıkları zaman bağımsız bakış açılarını kaybetmekte ve robot puanlama makineleri haline gelmektedirler. Linacre (2002a), her puanlamanın örtük özelliğe ilişkin performansın durumu ile ilgili bağımsız bilgi sağlaması gerektiğini söylemektedir. Karar vermede ise, puanların değil bu bilginin toplanması gerektiğini söylemektedir. KTK'nda puanlayıcılar arası güvenilirliği artırma çabalarının, gerçekte puanlayıcıların bağımsızlığını azalttığını ve bunun da ölçümlerin geçerliğini düşürdüğünü bulmuştur. Böylece, puanlayıcılar arası güvenilirlik zorlamayla artırıldığında, amaçlanan ölçümün güvenilirliğinin buna karşılık kurban edilmesi gerekmektedir.

Tüm bunlardan yola çıkarak, klasik test kuramının sınırlıklarını ele alırsak: 1) hatayı tek bir birim olarak ele almaktadır, 2) örneklem, testin amaçladığı popülasyonu temsil etmiyorsa güvenilirlik katsayısının kullanılabilirliği sınırlı olmaktadır (Hambleton, 1989) ve 3) bireyin puanı görevin (testin) zorluk/güçlük düzeyine göre yükselmekte veya düşmektedir (Downing, 2003; Downing, 2004; Hambleton, 1989).

KTK'ndaki güvenilirlik kestirimleri, hata varyanslarının puanlar üzerindeki etkilerini ele almadaki yetersizliğinden dolayı eksik kalmaktadır. Araştırmacıların farklı hata kaynaklarını nasıl daha doğru bir şekilde kestireceği ve geçerlik üzerindeki etkilerini nasıl kontrol edebileceği üzerine düşüncelerine yol açmıştır. Cronbach, Gleser, Nanda ve Rajartnam (1972) ölçme hatalarından kaynaklanan gözlenen puan ve gerçek puan arasındaki farklılıkları yönetmede iki yol önermiştir. Bunlardan birincisi, gözlenen puandaki hata kaynaklarının ifade edilmesidir. Diğer ise gözlenen puanı bir şekilde düzelterek gerçek puanın daha iyi bir kestiriminin sunulmasıdır. Cronbach'ın ilk yaklaşımını takiben, genellenebilirlik kuramı farklılıkları dikkate almıştır (Linacre, 1994): genellenebilirlik kuramı bireyleri tesadüfi bir şekilde gruplara atamakta ve "bireyin gözlenen ham puanının, benzer koşullar altında elde edilecek diğer ham puanlarla nasıl benzerlik gösterebileceğini" kestirmektedir. Bireyin testteki performansını farklı değişkenlik kaynaklarının (birey, görev, hakem vb.) bir fonksiyonu olarak kavramsallaştırmaktadır. Genellenebilirlik kuramı, bu değişkenlik kaynaklarının etkilerini ayırmakta ve bu etkilerin büyüklüğüne ilişkin istatistiksel kestirimler sunmaktadır. Bu kestirimler, varyans analizinde ana etki bireyler veya diğer değişkenlik kaynaklarından biri olduğu durumda, beklenen kareler ortalamasından elde edilen varyans bileşenleri şeklinde ifade edilmektedir. Buna rağmen, genellenebilirlik kuramı herhangi bir bireyin ham puanını belirli bir puanlayıcıya veya bireyin karşılaştığı göreve göre düzeltmemektedir. ÇDKRM ise Cronbach'ın ikinci yaklaşımını takip etmekte ve gözlenen puanları düzeltmektedir. ÇDKRM iki tane varsayıma dayanmaktadır: (a) bireyin yetenek düzeyi arttıkça, bireyin görevde başarılı olma olasılığı artmaktadır, (b) kolay bir görevde başarılı olma olasılığı, zor görevde başarılı olma olasılığından daha yüksektir. Linacre (2002b), ÇDKRM'nin her bir katılımcıyı bireysel olarak ele aldığını ve istatistiksel olarak mümkün olduğu sürece her bireyin ölçümünü, diğer bireylerin dağılımsal özelliklerinden, maddelerden ve puanlayıcılardan ayırmaktadır. Her bir

bireye ilişkin ölçümün hata dağılımı, ölçümlerin doğruluğunu ve istatistiksel geçerliğini araştırmada kullanılmaktadır. Test puanlarında çeşitli değişkenlik kaynaklarının etkisini kestirirken, ÇDKRM her bireyden elde edilen, doğrusal olmayan puanlama ölçeğine bağlı ham puanları, bireyin karşılaştığı puanlayıcılar ve görevler için düzelterek doğrusal ölçümlere dönüştürmektedir. ÇDKRM ile elde edilen sonuçlar standart hata ve uyum istatistikleri ile kontrol edilebilen, her bireyin performans düzeyine ilişkin doğrusal ölçümlerdir.

Hem genellenebilirlik yaklaşımı hem de ÇDKRM performans testleri geliştirmede, standart belirlemede, performansı analiz etmede ve puanlayıcı eğitiminde kullanılmaktadır (Kondo-Brown 2002; Kozaki, 2004; Lynch ve McNamara, 1998; Park, 2004; Weigle, 1998).

Puanlayıcı performansının değerlendirilmesinde, genellenebilirlik kuramı ve ÇDKRM gibi ölçme yaklaşımları, puanlayıcılar arası güvenilirliğin klasik kestiriminden daha fazla bilgi vermekte ve daha doğru dönütler sağlamaktadır. Stemler (2004)'in de belirttiği gibi puanlama ölçeğini nasıl uygulayacakları konusunda iki puanlayıcının anlaşmaya varmasına gerek yoktur, puanlayıcıların katılık düzeyindeki farklılıklar kestirilebilmekte ve her bir katılımcının son puanını hesaplamada dikkate alınmaktadır. Tabi ki, hakemlerin ölçülen yapıya ilişkin birbirine yakın bir anlayışa sahip olduklarından emin olabilmek adına bir miktar da uyum gereklidir. Puanlamanın doğru yapılıp yapılmadığı sadece iki hakem arasındaki uyum veya tutarlık derecesinin ne kadar yüksek olduğuna dayanmamalıdır, iki hakem arasındaki belirli bir miktardaki uyuşmazlık veya fikir ayrılığı sadece beklenen bir durum olarak nitelendirilmemelidir. Bu, aynı zamanda geçerlik için de önemlidir. Daha önemli bir adım, puanlayıcıların nasıl ve neden farklı olduklarını anlamak ve test puanlarında hakemin katılık düzeyinden kaynaklanan etkiyi telafi etmektir.

Bu çalışmanın kapsamı dâhilinde olan Genellenebilirlik kuramı ve ÇDKRM sonraki kısımda ayrıntılı bir şekilde ele alınmıştır.

1.4.2. Genellenebilirlik Kuramı

G-kuramı (Brennan, 2001; Cronbach, Gleser, Nanda ve Rajaratnam, 1972; Shavelson ve Webb, 1991), çeşitli değerlendirme ortamında, özellikle de performans değerlendirmelerinde artan bir şekilde kullanılmaktadır (Linn ve Burton, 1994; Shavelson, Baxter ve Gao, 1993) ve puanlayıcı değişkenliğini belirlemede KTK'ndan daha güçlü bir yaklaşım sunmaktadır. KTK'nda, bireyin gözlenen puanı, “gerçek puan” ve “hatadan” oluşmaktadır. Gerçek puanın varyansının gözlenen puan varyansına oranı ise güvenilirlik olarak tanımlanmaktadır (Crocker ve Algina, 1986). Buna rağmen, KTK'na bağlı olarak yapılan güvenilirlik kestirimleri sadece tek bir değişkenlik kaynağını dikkate almaktadır (Brown, 1991). G-Kuramı (Cronbach ve diğ., 1972) KTK'ı çerçevesini daha da genişleterek test puanlarına etki eden birden fazla değişkenlik kaynağını dikkate almaktadır. G kuramı, aynı anda farklı değişkenlik kaynaklarını belirleyebilmektedir. Ayrıca, bu kaynakların puanlama doğruluğuna etkisini kestirebilmekte ve araştırmacının çeşitli ve farklı uygulamaları denemesine olanak sağlamaktadır (Shavelson ve Webb, 1991). Yani, güvenilirliği değerlendirmenin yanı sıra güvenilir gözlemlerin tasarlanmasına, araştırılmasına ve kavramsallaştırılmasına da olanak sağlamaktadır (Brennan, 2001).

Klasik test kuramında oldukça önemli olan güvenilirlik, genellenebilirlik kuramında daha geniş ve daha esnek ele alınmıştır. Genellenebilirlik kuramı, gözlenen puanların gerçek puanları ne kadar doğru kestirdiği ile ilgilenmek yerine, gözlenen puanların bireyin davranışını tanımlanmış evrene genellemede ne kadar doğruluğa sahip olduğu ile ilgilenmektedir. Genellenebilirlik kuramı, gözlenen puanların evren puanını ne kadar iyi temsil ettiğini vermektedir (Shavelson, Webb ve Rowley, 1989).

Genellenebilirlik kuramı hata kaynaklarını değerlendirme açısından klasik test kuramından farklılık gösterir. Aynı ölçmeye ilişkin farklı kaynaklardan gelen hatalar güvenilirlik kestirimlerinde farklılığa yol açmaktadır. Ancak genellenebilirlik kuramında, hata terimi birçok değişkenlik kaynağından gelen hataları birlikte değerlendirmektedir (Brennan, 2001; Shavelson ve Webb, 1991).

Genellenebilirlik çalışmalarında, araştırmacı ölçme hatalarının farklı kaynaklarının görece önemini değerlendirebilmekte ve puanların güvenilirliğine ilişkin hem norm dayanaklı hem de ölçüt dayanaklı yorumlar yapabilmektedir. Genellenebilirlik kuramı, puan bağımlılığının ve ölçme hatalarının araştırılmasında eş zamanlı olarak birden fazla hata kaynağını analiz etmeye olanak sağlayan kapsamlı kavramsal bir çerçeve ve yöntem sunmaktadır (Brennan, 2001; Gao ve Brennan, 2001).

Genellenebilirlik kuramı, test-tekrar test güvenilirliği, iç tutarlık anlamında güvenilirlik, aynılık (benzer ölçekler) güvenilirliği (convergent) ve puanlayıcılar arası güvenilirliği birlikte değerlendirebilmektedir (Yin ve Shavelson, 2008). Ancak, KTK’nda güvenilirlik hesaplanırken, güvenilirliğin hangi amaç için kullanılacağı göz önüne alınarak, buna uygun bir yolla hesaplanması gerekmektedir (Baykul, 2000).

Shavelson ve Webb (1991)’e göre, genellenebilirlik kuramı klasik test kuramının daha genişletilmiş bir halidir. Genellenebilirlik kuramı, çoklu varyans kaynaklarını tek bir analizde ele alır ve her bir varyans kaynağının büyüklüğünün belirlenmesine olanak sağlar.

1.4.2.1. Genellenebilirlik Çalışması

Temeli varyans analizi üzerine kurulan genellenebilirlik kuramı, ölçme sonuçlarına karışan hataları, sadece bir değişkenlik kaynağından gelen hatalar olarak ele alan klasik kuramdan farklıdır (Shavelson ve Webb, 1991). Genellenebilirlik kuramı, puanlardaki değişkenliği farklı kaynaklara ve etkileşim terimlerine bölmede varyans analizini (ANOVA) kullanmaktadır. Birden fazla puanlayıcı tarafından bir grup bireye verilen puanlarda, bireyin gözlenen puanının toplam varyansı, değişkenliğin bağımsız kaynaklarına ayrılabilir ve bu kaynakların her birinin varyans bileşenleri kestirilebilir (Shavelson ve Webb, 1991). Varyans bileşenleri bir kez elde edildikten sonra, “bireyin gözlenen davranışlar örnekleme dayanan gözlenen puanından evren puanına ne kadar doğrulukla genelleme yapılabileceğini gösteren genellenebilirlik katsayısı hesaplanabilir” (Shavelson ve Webb, 1991).

G çalışmalarında hatanın her bir ana kaynağı (ör; hakemler, görevler) değişkenlik kaynağı (facet) olarak adlandırılmaktadırlar ve bu değişkenlik kaynaklarının her bir düzeyi (ör; hakem sayısı ve görev sayısı) koşul (condition) olarak bilinmektedir. G çalışmasının uygulanabilmesi için bu değişkenlik kaynakları ve koşulların tanımlanması gerekmektedir. (Shavelson ve Webb, 1991). Ölçme hatasına katkıda bulunan değişkenlik kaynakları sadece bireysel değişkenlik kaynaklarını içermemektedir, bunun yanında tüm değişkenlik kaynakları arasındaki etkileşimleri ve değişkenlik kaynakları ile ölçme nesnesi (ölçmenin amacı- object of measurement) arasındaki etkileşimleri de içermektedir (Shavelson ve Webb, 1991; Brennan, 2001; Yin ve Shavelson, 2008). G çalışmasının amacı, çok değişkenlik kaynağının olduğu bir evrende, her bir değişkenlik kaynağının varyansa katkısını araştırmaktır. Farklı kaynaklardaki gözlenen puanların varyansını tespit etmekte ve eldeki ölçme sonuçlarına göre bir güvenilirlik katsayısı da kestirmektedir (Brennan, 2001; Shavelson ve Webb, 1991).

Tek facetli bir desende, üç tane değişkenlik kaynağı bulunmaktadır: (a) ölçme nesnesi, (b) değişkenlik kaynağı (facet) ve (c) ölçme nesnesi, değişkenlik kaynağı ve artık veya tanımlanamayan varyans. Örnek; ölçme nesnesinin bireyler (p) ve görevlerin (t) ise değişkenlik kaynağı olduğu bir desende üç tane değişkenlik kaynağı olacaktır: p, t ve $p \times t$, e.

Farklı değişkenlik kaynağı desene girdikçe, olası etkileşimlerin sayısındaki artışa bağlı olarak değişkenlik kaynaklarının sayısı doğrusal olmayan bir şekilde artacaktır. İki facetli bir desende yedi tane değişkenlik kaynağı vardır. Eğer bu desendeki değişkenlik kaynaklarının görevler (t) ve durumlar (o) olduğunu düşünürsek, değişkenlik kaynaklarını şu şekilde sıralayabiliriz:

1. Bireyler (p)
2. Görevler (t)
3. Durumlar (o)
4. $p \times t$ etkileşimi
5. $p \times o$ etkileşimi

6. t x o etkileşimi

7. p x t x o, e etkileşimi, tanımlanamayan veya rastgele değişkenlik.

Üç değişkenlik kaynaklı ve tamamen çaprazlanmış bir desende değişkenlik kaynaklarının sayısı 15'i, dört değişkenlik kaynaklı çaprazlanmış desende ise 31'i bulunmaktadır (Brennan, 2001; Shavelson, Webb ve Rowley, 1989). Buz patenini ele aldığımızda s x t x d x r (sporcu, görev, boyut ve puanlayıcı) ve s x d x r (sporcu, boyut ve puanlayıcı) desenlerinin g kuramı analizlerinde değişken etkileri sırasıyla 15 ($\sigma^2_s, \sigma^2_t, \sigma^2_d, \sigma^2_r, \sigma^2_{st}, \sigma^2_{sd}, \sigma^2_{sr}, \sigma^2_{td}, \sigma^2_{tr}, \sigma^2_{dr}, \sigma^2_{std}, \sigma^2_{str}, \sigma^2_{sdr}, \sigma^2_{tdr}, \sigma^2_{stdr}$) ve 7 ($\sigma^2_s, \sigma^2_d, \sigma^2_r, \sigma^2_{sd}, \sigma^2_{sr}, \sigma^2_{dr}, \sigma^2_{sdr}$) tane ana ve etkileşimsel bileşeni içermektedir.

G- kuramı, performans değerlendirmede puanlayıcı değişkenliğini ve g çalışmasında yer alan diğer değişkenlik kaynaklarıyla çaprazlanan puanlayıcı değişkenliğini belirlemektedir. Birden fazla değişkenlik kaynağının olduğu performans değerlendirme desenlerinde ve analizlerinde G kuramı önemli bir araçtır. Spordaki yarışma ortamlarında (buz pateni, jimnastik, dalma) kullanılması tarafsız değerlendirmelerin yapıldığından emin olunmasına katkı sağlayacaktır.

G-kuramında değişkenlik kaynakları sabit veya değişken (fixed or random) olarak sınıflandırılmaktadırlar. Değişken facet, sayısız koşul ile ilişkilendirilmektedir. Eğer değişkenlik kaynaklarının seçilen düzeyleri rastgele örnekleme ile atanmışsa, bu aynı evrenden gelen bir başka örneklem ile değiştirilebileceğini göstermektedir. Bu durumda da değişkenlik kaynağı, değişken facet olarak sınıflandırılmaktadır. Örneğin; çarpma ile ilgili soruların değişkenlik kaynağı olarak alındığı bir araştırmada, o araştırma kapsamına alınan çarpma işlemleri, herhangi bir rastgele seçilen çarpma işlemleri kümesi ile yer değiştirebilir. Bu da değişkenlik kaynağının değişken olduğunu göstermektedir (Brennan, 1992; Brennan, 2001; Yin ve Shavelson, 2008).

Sabit değişkenlik kaynaklarında ise koşulların sayısı sınırlandırılmakta ve değiştirilmesi düşünülmemektedir. Örneğin; iki farklı konu alanını (ör; okuma ve yazma) değerlendirmek isteyen bir testteki maddeleri ele alalım, bu durumda araştırmacı değişkenlik kaynağını sabit olarak sınıflandırabilir. Çünkü sadece iki koşul bulunmakta

ve arařtırmacı diđer konu alanlarına genelleme yapmakla ilgilenmemektedir. Deđiřken ve sabit deđiřkenlik kaynakları arasında ayırım yapmak oldukça önemlidir. Çünkü bu seçim hatanın nasıl hesaplanacađını etkilemektedir (Shavelson ve Webb, 1991).

G- alıřmalarında, ölçme amacını (ölçme nesnesini) da kapsayan çeřitli deđiřkenlik kaynakları ile iliřkili varyanslar kestirilmektedir (Lynch ve McNamara, 1998). Deđiřkenlik kaynakları aynı zamanda tamamen aprazlanmış veya yuvalanmış olarak tanımlanabilirler. Bireyin deđerlendirmesinde çok sayıda puanlayıcı (r) ve görev (t) yer alıyorsa, en güçlü G alıřması deseni tamamen aprazlanmış desendir. Puanlayıcılar, görevler ve bireylerle aprazlanmıştır (p x t x r). aprazlanmış desenlerde, bir deđiřkenlik kaynađının her bir düzeyi diđer deđiřkenlik kaynađının her bir düzeyi ile birleřmektedir. Örneđin; eđer tüm puanlayıcılar tüm bireyleri tüm durumlarda puanlıyorsa, puanlayıcı ve durum deđiřkenlik kaynakları tamamen aprazlanmıştır. Bu řekilde tamamen aprazlanmış bir desende veriler bir kez toplandıktan sonra, puanların güvenilirliđi olası birçok deđiřik desende arařtırılabilir (Shavelson, Webb ve Rowley, 1989). Bu tür desenler, özellikle de büyük ölçekli performans deđerlendirmelerinde ender olarak tercih edilmektedir. Küçük görev grubu ve az sayıda bireyin olduđu ve tüm hakemlerin tüm kořulları puanladıđı özel durumlarda sıklıkla kullanılmaktadır. Ancak, bazen görev sayısı çok az olsa da, puanlanacak birey sayısı çok fazla olduđundan, puanlayıcıların tüm bireyleri puanlaması mümkün olmayabilir. Bu durumda, 1 ve 2. puanlayıcının ilk durumu, 3. ve 4. Puanlayıcının da ikinci durumu puanladıđı yuvalanmış desenler kullanılır (r:o) (Shavelson ve Webb, 1991). Hem aprazlanmış hem de yuvalanmış desenlerden oluřan desenler karma desen (mixed design) olarak adlandırılmaktadır. Puanlama iřlemi yuvalanmış desenleri ierdiđi zaman, yuvalanmış deđiřkenlik kaynađı ile iliřkili olan varyans, diđer deđiřkenlik kaynaklarından bađımsız olarak hesaplanamamaktadır. Buna örnek olarak; deđiřkenlik kaynaklarının puanlayıcılar (r) ve durumlar (o) olduđu iki deđiřkenlik kaynaklı tamamen aprazlanmış bir desende yedi tane deđiřkenlik kaynađı bulunmaktadır. Bunlar; bireyler (p), puanlayıcılar (r), durumlar (o), birey puanlayıcı etkileřimi (p x r), birey durum etkileřimi (p x o), puanlayıcı durum etkileřimi (r x o), ve birey, puanlayıcı, durum etkileřimi ve tanımlanmamış varyans (p x r x o, e). Ancak puanlayıcıların durumlarla yuvalanmış olduđu durumlarda, puanlayıcılar ile ilgili varyans, durumlarla ilgili

varyanstan ayırlanamamaktadır. Çaprazlanmış desendeki yedi değişkenlik kaynağına karşılık burada beş tane değişkenlik kaynağı bulunmaktadır. Bunlar ise; bireyler (p), durumlar (o), birey durum etkileşimi (p x o), puanlayıcı durum etkileşimi ile birleştirilmiş puanlayıcılar (r, rxo) ve birey, puanlayıcı ve durum etkileşimi ile birleştirilmiş birey puanlayıcı etkileşimi ve ölçülemeyen varyans (p x r, p x r x o, e). Tüm hakemler tüm durumları puanlamadığı için, sadece puanlayıcılara yüklenebilecek varyansın bilinmesi olanaksızdır. Böylece, puanlayıcı etkisi, durum puanlayıcı etkisi ile birleştirilmiştir. Puanlayıcılardan gelen varyans, durum puanlayıcı etkileşiminden ayırlamayacağından dolayı, birey puanlayıcı etkileşimine ilişkin varyansı da hesaplamak mümkün olmayacaktır. Birey puanlayıcı etkileşimi, üç yönlü etkileşim ile karışmakta ve ölçülemeyen hata olarak ortaya çıkmaktadır (Brennan, 1992; Brennan, 2001; Shavelson ve Webb, 1991; Yin ve Shavelson, 2008).

Büyük ölçekli performans değerlendirme durumlarında, iki puanlayıcı genellikle her bir bireyin seçkisiz atama yoluyla seçilen performans örneklemelerini değerlendirmektedirler. Aynı puanlayıcı çifti sistematik olarak bireylere bloklar şeklinde atanmamaktadır, fakat seçkisiz olarak her bir bireyi seçmektedirler. Bu durum, yuvalanmış desen olarak bilinen, hakemlerin bireylerle yuvalandığı desenden ((r:p)xt) farklıdır (Lee, Kantor ve Mollaun, 2002). Son yıllarda, araştırmacılar bu probleme pratik bir çözüm bulup, uygulamaya başlamışlardır. Puanlayıcılar (r) yerine, puanları (r') değişken facet olarak işleme almışlardır (Bachman, Lynch ve Mason, 1995; Lee, Golub-Smith, Payton ve Carey, 2001; Lee, Kantor ve Mollaun, 2002). Bu durumda, tüm bireylerin son puanı 2 puanlamadan elde ediliyorsa, iki değişken facet ile tamamen çaprazlanmış bir desen (p x t x r') kullanmak mümkündür.

G-kuramındaki bir diğer ayırım ise bağıl (relative) ve mutlak (absolute) kararlardır. Performanslara dayalı bağıl kararlar (G-katsayısı) ve mutlak kararlar (phi katsayısı) alınması için iki güvenilirlik katsayısı hesaplamaktadır. Bireyin diğer bireyler arasındaki durumu (norm referans) odak noktası olduğu zaman bağıl kararlar seçilmektedir. Örneğin; öğrencilerini performanslarına göre sıralamak isteyen bir orkestra şefi, bir öğrencisini diğer öğrenci ile karşılaştırırken öğrencinin kaç tane hata yaptığını önemseyecektir. Mutlak kararlar ise, performansın mutlak seviyesine göre bireyin ne

kadar iyi olduğuna dikkat edildiği durumlarda kullanılmaktadır, yani bireyin performansını değerlendirmede arkadaşlarının performansı önemsenmemektedir (ölçüt referanslı). Toplam hata varyansının nasıl hesaplanacağını bu ayırım belirlemektedir (Shavelson ve Webb, 1991).

1.4.2.2. G-Çalışmasından Elde Edilen İstatistikler

G-çalışmasının deseni tanımlandıktan sonra (ör; değişkenlik kaynaklarının sayısı, mutlak/görelî kararlar, yuvalanmış/çaprazlanmış desenler) puanlar arasındaki varyans bileşenlerine ayrılmakta ve dört olası istatistik elde edilmektedir. Varyansı ayırmada kullanılan istatistiksel model varyans analizidir (ANOVA). Her bir varyans bileşeninin büyüklüğü, her değişkenlik kaynağının toplam ölçme hatasına ne kadar katkıda bulunduğunu bize göstermektedir.

Dört tane istatistiğin iki tanesi ise görelî hata varyansı ve mutlak hata varyansıdır. Görelî hata varyansı görelî kararlar için kullanılırken, mutlak hata varyansı ise mutlak kararlar için kullanılmaktadır. Hata varyansları, G-çalışmasında kestirilen iki veya daha fazla varyans bileşeninin toplamıdır. Görelî kararlarda, hatayı tanımlamada ölçme nesnesi ile etkileşimde olan varyans bileşenleri kullanılmaktadır (Shavelson ve Webb, 1991).

Ölçme nesnesinin varyans bileşeni ve hata varyansları iki güvenilirlik katsayısı hesaplamada kullanılmaktadır. Görelî kararlar için hesaplanan güvenilirlik katsayısı g-katsayısı olarak bilinmekte ve paydasında görelî hata varyansını da içermektedir. Mutlak kararlarda ise kullanılan güvenilirlik katsayısı phi katsayısı (Φ) olarak bilinmekte ve paydası mutlak hata varyansını içermektedir (Brennan, 2001).

1.4.2.3. Karar Çalışmaları (K-Çalışmaları)

G-kuramı çerçevesinde araştırmacılar karar verici rolünü üstlenmektedirler. G-çalışması çeşitli hata kaynaklarının büyüklüğünü kestirirken, K-çalışmaları ise G-çalışmalarından

elde edilen bilgileri kullanarak, hata kaynaklarını en aza indirecek ölçme işlemlerinin tasarlanmasında kullanılmaktadır (Brennan, 2001; Shavelson ve Webb, 1991; Shavelson, Webb ve Rowley, 1989). G-çalışmalarının ilk kısmında kestirilen varyans bileşenleri araştırmacıya gözlenen puan ve evrendeki olası tüm puanlar arasındaki farkı açıklamaya olanak sağlamaktadır. G- çalışmasının ikinci kısmı ise, g çalışmasından elde edilen sonuçları araştırmacının uygun bir şekilde genellemesine olanak sağlayacak en uygun ölçme koşulu hakkında karar verilmesidir (Pedersen, Hagtvet ve Karterud, 2007). K-çalışmaları, puanlayıcı sayısı, puanlama durumu ve madde sayısı vb. değişkenlik kaynaklarından yapılan değişimlerin bir fonksiyonu olarak iki hata varyansının ve iki güvenilirlik katsayısının nasıl olacağını yansıtmaktadır. Buna ek olarak, K-çalışmaları, farklı desenler (ör; yuvalanmış desen) kullanılarak puanlama yapılırsa bu dört istatistikte meydana gelecek değişimleri yansıtmaktadır. Karar çalışmaları, çalışmaların gelecekte nasıl uygulanacağı ve nasıl daha iyi geliştirileceği konusunda bilgi vermektedir (Matt, 2010).

1.4.3. Madde Tepki Kuramı (MTK)

KTK, farklı yetenek düzeyindeki kişilerin belirli bir görevde nasıl performans göstereceğine ilişkin bilgi vermemektedir. Madde tepki kuramı, temelde bireyin gözlenemeyen yetenekleri ile ölçme aracındaki maddeler arasındaki ilişkiyi matematiksel bağıntılarla açıklamaya çalışan bir kuram olarak geliştirilmiştir. Yani, madde-tepki kuramı, ölçülen özellik bakımından farklı yetenek düzeylerindeki cevaplayıcıların maddeyi nasıl yanıtladıklarına ilişkin matematiksel bir model sunmaktadır (Crocker ve Algina, 1986). MTK, bize test durumunda maddenin fonksiyonunun daha detaylı resmini vermekte ve bireyin gerçek yeteneğini anlamamıza olanak sağlamaktadır. MTK ölçme modeli araştırmacının veri ve model arasındaki uyumsuzlukları ortaya çıkarmasına ve farklı yetenek düzeylerinde yer alan bireyleri birbirinden ayırt edemeyen maddelere ilişkin sorunları belirlemesine olanak sağlamaktadır. Ölçülen özellik açısından bireyin yeteneğini en iyi yansıtacak şekilde ölçme aracının düzenlenmesine olanak sağlamaktadır (Hambleton ve Swaminathan, 1985).

Madde tepki kuramıyla ilgili yürütülen ilk çalışmalarda bir bireyin bir testte yer alan maddelere verdiği cevapların doğru veya yanlış olarak sınıflandığı modeller üzerinde durulmuştur. Bir, iki ve üç parametrelili madde tepki kuramı modelleri, iki kategorili (0-1) modeller olarak bilinirler (De Ayala, 1993). Madde tepki kuramında bir soruya doğru cevap verme olasılığının yetenek düzeylerine göre dağılımını gösteren dağılım fonksiyonu madde karakteristik eğrisi olarak adlandırılmakta, bu eğriyi tanımlayan farklı matematiksel denklemler de modelleri oluşturmaktadır.

MTK birçok durumda yaygın olarak kullanılabilir. Bir parametrelili model madde güçlüğü parametresi (bi) ve bireyin yetenek düzeyi arasında ilişki kuran lojistik bir modeldir. Kestirilen madde parametrelerinin sayısı MTK modelini tanımlamada kullanılmaktadır. Rasch modeli bireyin yeteneğini ve madde özelliklerinden sadece bir tanesini (uyum eğilimi – agreement tendency veya madde güçlüğü) kestirmektedir. Bu nedenle bir parametrelili model olarak bilinmektedir. Rasch, nitelikli gözlemleri doğrusal ölçümlere dönüştürmenin ve belirli bir kategoride gözlenme olasılığı ile bireyin yeteneği ve madde güçlüğü arasındaki ilişkilerin incelenmesinin önde gelen isimlerindendir. Rasch modeli iki değişkenin (yetenek ve madde güçlüğü) ölçümünde tek boyutu tanımlamada kullanılmasına rağmen performans değerlendirmede de kullanılması uygundur. Bunun nedeni dereceleme (puanlama) ölçeği kullanan hakemlerden elde edilen veriyi işlevsel (etkili) bir şekilde ele alabilmektedir (Hambleton, Swaminathan ve Rogers, 1991).

Rasch modelinin esnek ve güçlü yapısı seçim (uygun bireylerin seçilmesi) yapmaya oldukça uygun görülmektedir. Rasch modelinin ilkeleri bu gibi durumlarda avantaj sağlamaktadır; çünkü tüm bireylerin, tüm koşullarda tüm puanlayıcılar tarafından puanlanması mümkün olmayabilir (Fox ve Jones, 1998).

Rasch modeli, araştırmacılara diğer modellerin veremeyeceği bilgiler sağlamak ve diğer modellerle gözlenemeyecek sorunların belirlenmesine olanak sağlamaktadır. Varyans analizi (ANOVA) bir takım analizlerde kullanılabilir ancak varyans analizi eksik verilerin ve standart olmayan ham puanların düzeltilmesinde zorluklar yaşamalarına yol açabilir (Lunz, Wright ve Linacre, 1990). Linacre (1997) tam olmayan

veri setlerinin bile Rasch modeli için sorun yaratmadığını çünkü genellikle en iyi kestirimin hesaplanabileceğini ileri sürmektedir. Bu durum da Rasch modelinin beşeri bilimlerde, özellikle de eğitim ve sağlık bilimlerinde kullanılabilecek en iyi ölçme modellerinden biri olduğunu göstermektedir.

Bir parametrelili model madde güçlüğü parametresi (b) ve bireyin yetenek düzeyi arasında ilişki kurmaktadır. İki parametrelili modelde ise madde güçlük parametresinin yanına ayırt edicilik parametresi (belirli bir yetenek düzeyinde sorunun düşük ve yüksek yetenek grubundakileri ayırma gücü – a) eklenmektedir. Lojistik modeller içerisinde madde karakteristik eğrisini üç ayrı parametre ile tanımlayan model üç parametrelili lojistik modeldir ve bu modelde; madde ayırtediciliği (a), madde güçlüğü (b) ve tahminle doğru cevabı bulma olasılığı, yani şans faktörü parametresi (en düşük (minimum) yetenek düzeyindeki bireylerin maddeyi doğru yanıtlama olasılığı – c) olmak üzere üç parametre söz konusudur (Baker, 1987; Hambleton ve Swaminathan, 1985).

Andrich iki seçenekli maddeler için geliştirilen Rasch modelinin ilkelerini kullanarak, dereceli puanlama ölçeği modelini geliştirmiştir. Bu model, ikiden fazla cevap seçeneği bulunan maddeleri analiz edebilmek için iki seçenekli maddeler için kullanılan Rasch modelinin gelişmiş şeklidir. Bu modelin en yaygın kullanımı, seçenekler arasındaki mesafenin aynı olduğu Likert tipi ölçeklerdedir (tamamen katılıyorum, katılıyorum, kararsızım, katılmıyorum, tamamen katılmıyorum) (Linacre, 2007).

Masters (1982), Andrich'in yaptıklarının da ötesine giderek, kısmi doğruluğu gösteren ve her bir maddeye özgü olan kısmi puanlama modelini geliştirmiştir. Masters'a (1982) göre, kısmi puanlama modelini geliştirmenin altında yatan mantık, bireyin genel yeteneği ile ilgili mümkün olduğunca daha fazla bilgi sağlamaktadır. Sonuca ulaşmak için 5 basamakta çözülebilen bir matematik problemi düşünelim, birey basit bir hesaplama hatası dışında tüm basamakları doğru yapmasına rağmen yanlış bir sonuca ulaşmıştır. Eğer soru, 1–0 şeklinde (iki seçenekli model) puanlansaydı, birey bu sorudan 0 alacaktı çünkü doğru sonucu bulamamıştır. Ancak kısmi puanlama modeli ile birey

4/5 olabilir, bu da onun doğru sonuca çok yakın olduğunu göstermektedir. Bu modelde, her bir maddenin kestirim sayısı kategori sayısına göre değişebilir.

1.4.3.1. Çok Değişkenlik Kaynaklı Rasch Modeli

Yapılan çalışmalar çok değişkenlik kaynaklı Rasch modelinin çeşitli durumlara başarılı bir şekilde uygulanabileceğini göstermiştir (Chang ve Chan, 1995; Engelhard, 1992; Smith ve Kulikowich, 2004). ÇDKRM, orijinal Rasch ölçme modelinin geliştirilmiş halidir. Bireyin yeteneği ve madde gücünün yanında test durumları ile etkileşen diğer faktörleri de ölçme modeline almaktadır. Sadece tek bir değişkenlik kaynağını değerlendirmek yerine, ÇDKRM puanlayıcılardan, durumlardan ve görevlerden gelen sistematik hata kaynaklarını eş zamanlı olarak içeren birden çok değişkenlik kaynağını değerlendirebilmektedir (Sudweeks, Reeve ve Bradshaw, 2005). ÇDKRM, Linacre tarafından geliştirilmiştir (Bond ve Fox, 2007). Üç değişkenlik kaynaklı Rasch modeli aşağıdaki gibi gösterilebilir:

$$\text{Log}_e (P_{nij}/P_{nij(k-1)}) = B_n - D_i - C_j - F_k$$

P_{nij} : n bireyinin i maddesinden, j puanlayıcısı tarafından k puanı alma olasılığı

$P_{nij(k-1)}$: n bireyinin i maddesinden, j puanlayıcısı tarafından k-1 puanı alma olasılığı

B_n : n bireyinin performans ölçümü

D_i : i maddesinin güçlük düzeyi

C_j : Puanlayıcının katılığı

F_k : k basamağında (kategorisinde) puanlamanın, k-1 basamağına (kategorisine) göre zorluğu (güçlüğü) (Lunz, Wright ve Linacre, 1990).

Çok değişkenlik kaynaklı Rasch modeli bileşenlerin veya değişkenlik kaynaklarının etkileşimlerine ilişkin sonuçlardan elde edilen bilgilerle oluşturulmuştur. Değişkenlik kaynakları ölçme işlemlerinin “durumları” olarak düşünülmektedir. Bunlar; puanlayıcılar, test maddeleri, görevler, bireyin yetenekleri, ölçme seansları, ölçme koşulları vb. olabilir. Performans değerlendirmede genellikle ölçme denklemine eklenen

değişkenlik kaynağı hakemler tarafından puanlanan maddelerdir. Linacre (1994) “Facets” adlı bilgisayar programını geliştirmiş ve bunu: “ÇDKRM, Rasch modelinin ikiden fazla değişkenlik kaynağı bulunan durumlara genişletilmiş şekli” olarak açıklamıştır.

ÇDKRM, G-kuramı gibi hata varyansını birçok kaynağa ayırabilmesinin yanında bunu daha fazla parçaya da bölebilmektedir. ÇDKRM, G-kuramında hesaplanan ana etkiye benzer olarak her bir değişkenlik kaynağı için grup düzeyinde istatistikler vermektedir, bunun yanında bireysel düzeyde de istatistikler sağlamaktadır. Araştırmacının her bir bireyi, hakemi, durumu ve maddeyi değerlendirmesine olanak sağlamaktadır. Eğer bir puanlayıcı diğer puanlayıcılardan daha katı puanlama yapıyorsa, araştırmacı ÇDKRM’ni kullanarak bunu belirleyebilir ve buna müdahale edebilir. ÇDKRM araştırmacıya her bir değişkenlik kaynağı ile ilgili daha detaylı bilgi sunmaktadır. Myford ve Wolfe (2003) yaptıkları araştırmada, ÇDKRM’nin beş puanlayıcı hatasının belirlenmesine olanak sağladığını da belirtmişlerdir. Bu puanlayıcı hataları: (a) cömertlik/katılık, (b) merkezi eğilim, (c) ranj kısıtlaması, (d) halo ve (e) ayrımcı cömertliktir (differential leniency).

ÇDKRM performans değerlendirme sonuçlarını analiz etmede aşağıdaki avantajları sunmaktadır: (a) ölçme ortamında bir veya daha fazla ek bileşeni (faceti) kapsayacak şekilde Rasch modelini geliştirmektedir, (b) bireyin yeteneği, görevin güçlüğü ve puanlayıcıların katılığını belirlemede ham puanlar yeterli istatistiklerdir, (c) puanlayıcıların puanlamasını kestirmede her kategorideki gözlemlerin sayılması yeterli istatistiklerdir, (d) verilerin modele uyumunu test etmede kullanılan uyum istatistikleri kalite kontrolünü desteklemektedir, (e) uyumsuzluğun birçok şekline karşı güçlüdür/dirençlidir, (f) eksik verilere karşı güçlüdür, eksik verilerin tamamlanması gibi bir gereklilik yoktur, (g) kestirim yaparken, model verilerdeki tesadüflük (randomness) uygunmuş (well behaved) gibi işlem yapmakta ve (h) veriden yapılandırılan ölçümler doğrusal referans çerçevesinde gözlenmektedir (Linacre, 2001). Bu son özellik Facets bilgisayar programının çıktılarından yararlanacak olan kullanıcılar için oldukça önemlidir, oluşturulan grafiklerin (değişken haritaları) geometrik özellikleri kullanıcıların anlayabileceği şekildedir. Facets bilgisayar programında analiz yapıldığı

zaman tüm deęişkenlik kaynakları eş zamanlı ama birbirinden bağımsız olarak analiz edilmekte ve tek bir doğrusal ölçekte (logit ölçek) kalibre edilmektedirler. Deęişkenlik kaynaklarının birlikte kalibrasyonu, hakem katılığına ilişkin ölçümün, bireyin performansı ve görev güçlüğü ile aynı ölçekte ifade edilmesini olanaklı kılmaktadır. Bu yaklaşımın bir avantajı da her bir bireyin her maddede her hakem tarafından puanlanmasını gerektirmemesidir. Buna rağmen, hakemler bireyleri puanlarken araştırma deseninin dikkatli şekilde planlanması gerekmektedir. Bu plan, her deęişkenlik kaynağındaki her bir gözlemi, aynı deęişkenlik kaynağındaki diğer gözlemlerle doğrudan veya doğrudan olmayan bir biçimde ilişkilendirecek bir ağ yaratmalıdır. Her bir hakem, geriye kalan tüm hakemlerle doğrudan veya dolaylı bir şekilde bağlantılı olmalıdır. Böylelikle araştırmacı çeşitli deęişkenlik kaynakları arasında bilgi sağlayacak karşılaştırmalar yapabilecektir (Linacre ve Wright, 2002). Her deęişkenlik kaynağındaki her gözlem için Facets analizi sonucunda bir ölçüm (kalibrasyonun logit kestirimi), standart hata (bu logit kestirimin doğruluğu hakkında bilgi vermektedir) ve uyum indeksleri (deęişkenlik kaynağı ile ilgili bilginin, ölçme modelinin beklentileri ile ne kadar iyi uyum sağladığına ilişkin bilgi vermektedir) elde edilmektedir. Bu bireysel düzeydeki istatistiksel bilgilere ek olarak, grup düzeyinde de istatistiksel bilgiler sağlamaktadır. Bu da verinin genel örüntüsü hakkında yararlı bilgiler vermektedir. Örneğin ki-kare testi, ayırım istatistięi, özet uyum istatistikleri (Wright ve Linacre, 1994; Myford ve Wolfe, 2003).

Ki-kare testi, hakemlerin katılıkları arasında anlamlı bir fark bulunmamaktadır hipotezini test etmektedir. Ki-kare deęeri, en az iki hakemin katılık konusunda farklı olduğunu göstermektedir. Bu durumda, test puanlarının yorumlanması ve test puanlarının geçerliğinde hakemlerin katılıklar arasındaki farkı dikkate almak gerekmektedir (Wright ve Linacre, 1994).

ÇDKRM, aynı zamanda bize uyum istatistikleri vermektedir. Bu uyum istatistikleri, grup veya bireysel düzeyde her bir deęişkenlik kaynağının, ÇDKRM tarafından kestirilen beklenen deęerlerle ne kadar iyi uyum gösterdiğine ilişkin bilgi vermektedir. Uyum istatistikleri, ki-kare deęerinin serbestlik derecesine bölünerek hesaplanan kareler ortalaması şeklinde rapor edilmektedir. Uyum istatistikleri, iç uyum (infit) ve dış uyum

(outfit) istatistiklerini içermektedir. Dış uyum istatistikleri, aşırı değerlerden (outliers) oldukça fazla etkilenmektedir. Diğer yandan, iç uyum istatistikleri ağırlıklandırılmıştır ve küçük artıklardaki veya aşırı olmayan değerlerdeki beklenmeyen örüntülere daha duyarlıdır. Uyum istatistiklerinin beklenen değeri 1'dir ve 0 ile sonsuz arasında değerler almaktadır. Eğer uyum istatistiği 1'den küçükse, veri gereksiz, bağımlı veya kısıtlıdır. Buna rağmen, eğer uyum istatistiği 1'den büyükse, bu verilerin tutarsız olduğunu, beklenmeyen değişkenlik içerdiğini veya aşırı uçlarda olduğunu göstermektedir (Smith ve Kulikowich, 2004).

Eğer değişkenlik kaynağı beklenen performansı gerçekleştirmişse, hem iç uyum hem de dış uyum istatistikleri 0.5 ile 1.5 (veya amaca bağlı olarak 0.4 ile 1.2) arasında yer alacaktır (Linacre, 2002a). G-kuramının sağladığı genel güvenilirlik istatistiği olarak bilinen g-katsayısı gibi, ÇDKRM de iki ayrı güvenilirlik istatistiği vermektedir. Bunlar ayırt etme indeksinin (separation index) güvenilirliği ve ayırt etme oranıdır (separation ratio). Ayırt etme indeksinin güvenilirliği, iç tutarlığın kestirimi olan Cronbach alfa katsayısı ile benzerdir ve 0 ile 1 arasında değer almaktadır. Bu istatistik, değişkenlik kaynağının içindeki koşullar veya öğeler arasında ne kadar varyans olduğunu bir doğru boyunca göstermektedir. Varyansın bireyler arası gerçek farklılıklardan gelmesi arzu edilmektedir. Birey için ayırt etme indeksi güvenilirliğinin 1'e olabildiğince yakın olması ve diğer tüm değişkenlik kaynaklarında ise 0'a olabildiğince yakın olması istenen bir durumdur. Ayırt etme oranı ise 0 ile sonsuz arasında değerler almaktadır. Ayırt etme indeksinin güvenilirliği gibi, ayırt etme oranının da birey değişkenlik kaynağında olabildiğince yüksek, diğer değişkenlik kaynaklarında ise olabildiğince düşük olması arzu edilmektedir (Wright ve Linacre, 1994; 2001).

Çok değişkenlik kaynaklı Rasch modelinin insan performansını değerlendirmede diğer ölçme modellerine göre temel avantajı ölçümlerin aralık ölçeğinin log oranı ile yapılmasıdır (a log odds scale of interval unit). Uzunluk veya ağırlık ölçümleri ile karşılaştırıldığı zaman, beşeri bilimlerde kullanılan birçok ölçekten elde edilen ölçüm birimleri eşit aralıklarda temsil edilmemektedir. Buna rağmen, bu birimler ÇDKRM'nde eşitlenmiş aralıklara dönüştürülmektedirler (Linacre, 2002a).

Bond ve Fox (2007) insan performansının sıralama ölçeğinden, log orana nasıl dönüştürülebileceğini ve bu logit ölçeklerin insan performansını anlamada niye daha fazla bilgi verdiğini açıklamıştır. Bond ve Fox'a göre, ham puanların veya yüzdelerin kullanımı öğrencileri ortalama puanlar etrafında toplamaya eğilimlidir ve bu da yetenekli ve daha az yetenekli olan bireylerin sonuçlarının uygun olmayan, yetersiz bir karşılaştırmasını sunmaktadır. Test sonuçlarının orta noktasına yakın bir yerlerde birkaç fazladan puan kazanmak, örneğin 48'den 55'e çıkmak için gerekli olan yetenek artışı ile 88'den 95'e veya 8'den 15'e çıkmak için gerekli olan yetenek artışı aynı değildir. Buradaki esas problem kesir veya yüzde puanlar arasındaki mesafelere doğrudan anlamlar vererek, bu verilerden bireylerin veya maddelerin sıralaması ile ilgili çıkarımlar yaparken hatalar yapılmasıdır.

ÇDKRM'nde performans görevinde başarılı olma şansı, performans ortamında yer alan değişkenlik kaynaklarının sayısına göre belirlenmektedir. Bu değişkenlik kaynakları, bireyin yeteneğini, görevin güçlük düzeyini, puanlayıcının özelliklerini ve performansın gösterildiği ve puanlandığı performans ortamının özelliklerini içerebilmektedir. Bu değişkenlik kaynaklarının birbirleri ile ilişkileri, bireyin belirli bir yetenek düzeyinde belirli bir görevden belirli bir puanı alma olasılığını artırmakta veya azaltmaktadır (Lumley ve McNamara, 1995).

Rasch modeli insan performansını belirlemeye oldukça uygundur, çünkü (a) incelenen yeteneklerin kazanımını belirlemeye duyarlıdır, ör; gelişim veya kazanım sırasını, (b) yetenekler veya bireyler arasındaki gelişimsel farkları kestirebilir ve bir bireyin diğer bireyden ne kadar ileride olduğunu gösterebilir, (c) maddeler ve bireylerin gösterdiği genel gelişim örüntülerinin her bir madde ve her bir bireye ilişkin gelişim örüntüleri ile uyumlu olup olmadığını belirlemeye izin vermektedir (Linacre, 2002a; 2002b).

Linacre (1989) ÇDKRM'nin performans değerlendirilmesinde kullanımının avantajlarını açıklamıştır. Model-veri uyumu sağlandığı zaman ÇDKRM avantajları (a) her bir parametre değerleri tüm diğer parametrelerden bağımsızdır, (b) parametreler birleştirilmekte ve tek bir doğrusal ölçekte yer almaktadır, (c) parametre kestirimi, tüm puanlamaların bir araya toplanmasına bağlı olmasına karşın, bu puanlamaların herhangi

birinin değerlerinden bağımsızdır (Linacre, 1989). Bu da, bireyin ölçülen yeteneği olarak yorumlanmakta ve maddelerin değerlendirilmesine bağlı olmamaktadır ve maddenin güçlüğünün değerlendirilmesi ise performans değerlendirmedeki bireylere bağlı değildir.

Puanlayıcı değişkenlik kaynağındaki uyum istatistikleri ise gözlenen puanların, model tarafından belirlenen beklenen puanlardan ne kadar sapma gösterdiğine ilişkin bilgi vermektedir. Beklenen puanlar, Rasch ölçme modelinin belirlediği ve hakemin o maddede bireye vereceği puandır. Bu puanlar belirlenirken hakemin tüm uygulamalardaki katılık derecesi, o maddede bireye diğer hakemler tarafından verilen puanlar, maddenin kestirilen güçlük düzeyi ve bireyin kestirilen yetenek düzeyi dikkate alınmaktadır. Beklenen ve gözlenen puanlama arasındaki büyük farklar, hakem etkisinin varlığını göstermektedir. Uyumsuzluk, gözlenen ve beklenen değerler arasındaki büyük farklılığı göstermektedir. Bu da araştırmacıların hakemin o gözlemdeki puanlama performansına dikkat etmeleri gerektiğini göstermektedir. Rasch modelinin tek boyutluluk varsayımını karşılamak için, değişkenlik kaynaklarının kestirimleri analiz sırasında tek boyutlu ve doğrusal olmaya zorlanmaktadır. Belirli bir gözlemdeki uyumsuzluk, verinin beklenen kestirimle eşleşmediğini göstermektedir, fakat bu Rasch kestirimlerinin doğrusal olmadığını göstermemektedir (Bond ve Fox, 2007).

Sonuç olarak, ÇDKRM performans değerlendirmedeki puanlayıcı etkisini açıklamada, puanlayıcı performansını kontrol etmek için kullanılan diğer yaklaşımlarla karşılaştırıldığında avantajlı görünmektedir. Puanlayıcı performansı ile ilgili daha doğru kestirimler yapmakta ve bireysel düzeyde puanlayıcı davranışı ile ilgili daha ayrıntılı bilgi vermektedir. Bunun yanında puanlayıcı eğitimini sürdürmek için de zengin bilgi sağlamaktadır. Bu da puanlayıcının daha doğru değerlendirme yapmasına ve test puanlarının geçerlik ve güvenilirliğini artırmaya katkı sağlamaktadır. Gözlemcilerin, hakemlerin ve puanlayıcıların gözlemlerine dayalı değerlendirme yapılan çalışmalarda çok değişkenlik kaynaklı Rasch modelinin kullanımı oldukça başarılıdır (Engelhard, 1994; Lunz, Wright ve Linacre, 1990; O'Neill, 1999).

1.4.4. Genellenabilirlik Kuramı ve Çok Değişkenlik Kaynaklı Rasch Modelinin Karşılaştırılması

G-kuramının ve ÇDKRM birlikte kullanımı 2004–2005 yıllarında popüler olmuştur. Birçok araştırmacı, hangisinin avantajlı olduğu sorusuna cevap bulabilmek, benzerlik ve farklılıklarını ortaya koyabilmek için bu yöntemler arasındaki ilişkiyi araştırmışlardır.

ÇDKRM’nde bireyin yetenek kestirimi “ölçüm/measure” olarak adlandırılmaktadır. Böylece bireyin orijinal ham puanları farklılaşmaktadır. Orijinal puanlama, sıralama ölçeğinin özelliklerini taşımaktadır ancak Rasch modelinde veri uyumu sağlandığında, ölçümler eşit aralıklı ölçek özelliğini taşımaktadır (Embretson ve Reise, 2000). Bu da ÇDKRM’ni kullanmanın bir avantajıdır. G-kuramında bu elde edilmemektedir. G-kuramında her birey için gözlenen ortalamalar bireyin yeteneğinin bir kestirimidir ancak bu ortalamalar orijinal puanlarla aynı ölçektir. Genellenebilirlik kuramı, ölçek gerçekte sıralı olsa bile aralık ölçeğindeki gibi işlem yapmaktadır. Bu da bireyler veya maddeler arasındaki geçerli karşılaştırmaları zora sokmaktadır. Puan çiftleri arasındaki sıra puanlarındaki eşit farklılıklar, incelediğimiz yapının eşit miktarlarını göstermemektedir (Smith ve Kulikowich, 2004).

ÇDKRM ölçmenin standartlaştırılmış birimlerine (lojit veya logaritma) dayandığından dolayı birkaç değişkenlik kaynağını birbiri ile karşılaştırma imkânı vermektedir. Yetenek ölçümleri her bir öğrenci için hesaplanmakta ve görev gücüyle hakemlerin katılımındaki farklılıklar düzeltilmektedir. Bireyin yetenek düzeyi görevlerin ya da hakemlerin dağılım özelliklerinden etkilenmemektedir. Sonuç olarak, bir birey zor bir görevde katı bir puanlayıcı tarafından puanlansa bile, bireyin yetenek puanı örneklemdaki diğer bireylerle doğru bir şekilde karşılaştırılabilir (Embretson ve Reise, 2000; Smith ve Kulikowich, 2004).

ÇDKRM örneklemin dağılımına bağlı değildir. Aynı popülasyondan seçilen çeşitli örneklerde değişkenlik kaynağının kestirimleri aynı kalmaktadır. Genellenebilirlik kuramı bu özelliği taşımamakta ve örneklemin dağılımına bağlı olmaktadır. Genellenebilirlik çalışmalarında, istatistikler puanlayıcıların katılımından ve görevin

zorluğundan etkilenmektedir. G-kuramı maddeler ve hakemler arası homojenliği gerektirmektedir. Her bir hakem veya görev olasılıklar evrenindeki herhangi bir hakem veya madde ile değiştirilebilmelidir. ÇDKRM değişkenlik kaynakları içerisinde heterojenliği desteklemektedir (Smith ve Kulikowich, 2004).

G-kuramı her bir değişkenlik kaynağının toplam varyansa katkısını vermektedir. Hata miktarını görmemize yardımcı olmaktadır. ÇDKRM ise puanlayıcılar arasındaki değişkenliğin çok fazla olduğunun belirlenmesine olanak sağlamasına rağmen puanlayıcılara ilişkin hataların diğer değişkenlik kaynaklarına göre ölçme hatalarına ne kadar katkıda bulunduğunu göstermemektedir. Her bir değişkenlik kaynağının ölçme hatalarına katkılarını ve görece büyüklüklerini anlamak, güvenilirliği artırmada hangi değişkenlik kaynağının ilk olarak dikkate alınması gerektiğini belirleme konusunda yardımcı olmaktadır. ÇDKRM maddeler, bireyler ve puanlayıcılarla ilgili hem grup düzeyinde hem de tek tek görevler, bireyler ve puanlayıcılar ile ilgili ayrıntılı bilgi verirken, genellenebilirlik kuramının grup düzeyinde bilgi verdiğini ortaya koymuşlardır (Smith ve Kulikowich, 2004; Ure, 2011).

Genellenebilirlik kuramının bir diğer avantajı ise karar çalışmalarının uygulanmasını olanaklı kılmasıdır. ÇDKRM'nin böyle bir işlevi yoktur. Dengelenmemiş çalışma (unbalanced) desenleri ve eksik verilerin G-kuramı ile analiz edilmesi zordur. G-kuramı bu gibi durumları etkili bir şekilde ele alamamaktadır. ÇDKRM ise veri kümesindeki bu bozuklukları genellenebilirlik kuramına göre daha iyi bir şekilde ele alabilmektedir (Ure, 2011).

Genel olarak, yukarıda da tartışıldığı gibi, iki kuramın da avantajları bulunmaktadır. Ancak, iki model arasındaki bu farklılıklardan dolayı Smith ve Kulikowich (2004) araştırmacıların amaçlarına uygun olan modeli seçmelerini tavsiye etmektedirler. Linacre ve Wright (2004) genellenebilirlik kuramı ve ÇDKRM aynı problemi çözmeye çalışan iki rakip teknik olarak görülmesine rağmen bunun söz konusu olmadığını savunmaktadır. İki tekniğin kullanıldığı veri setleri benzer olsa da amaçlarının tamamen farklı olduğunu söylemektedirler. G-kuramı değişkenliği gelecekteki veri toplama süreci

için düzeltmeyi amaçlarken, ÇDKRM bunu çalışılan veri kümesi üzerinde yapmayı amaçlamaktadır, çalışılan verilerdeki değişkenliği düzeltmeye çalışmaktadır.

Her iki tekniğin farklı avantajlarının olmasından dolayı, bir tekniğin diğerine tercih edileceği durumla karşılaşıldığı zaman, hangi tekniğin kullanılacağına karar vermede araştırma probleminin dikkate alınması gerekmektedir. Kesin ve doğru ölçümler yapabilmek için doğru araçlara ihtiyaç duyulmaktadır, Rasch Modeli ve genellenebilirlik kuramı da birçok ölçme modeli gibi değişkenlik kaynaklarının ölçümünde bireylerin taleplerini ve ihtiyaçlarını karşılamak için geliştirilmiştir. Farklı ölçme tekniklerini bir arada kullanmak, araştırmacıların daha bütüncül daha kapsamlı yaklaşımlar elde edebilmesinin ve ölçmek istediklerinin daha doğru temsillerini sunabilmelerinin bir yolu olarak görülmektedir (Sebok, 2010).

1.5. PROBLEM DURUMU

Eğitimde ve psikolojide olduğu gibi spor bilimlerinde de performansın ölçülmesi ve değerlendirilmesi önemli bir yer tutmaktadır. Performansın ölçülmesinde kullanılan yöntemler; davranışın gözlenip ölçülmesi, sonuca bakarak bireyin yetisi hakkında bir değer yargısına varılması ve çeşitli basamaklarda bireyin gözlenip, ara ara puanlar verilmesidir (Turgut, 1993). Sporda yapılan ölçmelerde hakemlerin ölçme sürecine katıldığı spor dalları bulunmaktadır. Hakemlerin ölçme sürecine katıldığı durumlarda güvenilirlik çalışmaları önem kazanmaktadır. Spor yarışmalarında, sporcuların hakemler tarafından öznel olarak değerlendirilmesi spor izleyicisinin kafasında soru işareti bırakmaktadır. Hakemlerin, subjektif değerlendirmeleri sonucunda ne kadar doğru, adil, geçerli ve güvenilir kararlar verilmekte olduğu spor performansının ölçülmesinde önemli bir sorun olarak karşımıza çıkmaktadır. Birden fazla hakemin görüşünden oluşan bu değerlendirmeler öznel kararlara karşı bir tedbir olsa da güvenilirlik çalışmaları için birden fazla değişkenlik kaynağının birlikte alınması gerekliliğini ortaya çıkarmaktadır. Tüm değişkenlik kaynaklarının birlikte göz önüne alındığı kapsamlı bir güvenilirlik katsayısının elde edilmesinde genellenebilirlik kuramı ve çok değişkenlik kaynaklı Rasch modeli kullanılabilir. Bunun yanında, hakemlerin katılık ve cömertliğinin,

hakemler arası uyumun yanında hakemim kendi içindeki uyumunun ve hakemlerin yanlılık düzeylerinin belirlenmesi önem taşımaktadır. Görevlerin güçlüğü ve birbiri yerine kullanımının uygunluğu ve yapılan yetenek kestiriminin (sporcuların sıralanmasının) uygunluğunun belirlenmesi de performans ölçümlerinde dikkat edilecek noktalar olarak görülmektedir. Ölçme işlemindeki sistematik ve tesadüfi hataların ayrılmasına duyulan ihtiyacın yapılacak analizlerle karşılanabileceği düşünülmektedir.

Artistik buz pateni hem bir sanat hem de bir spor dalı olarak, teknik yetenekler ve artistik alanlarda değerlendirilmektedir. Popüler ve olimpiik bir spor dalı olan buz pateni, tekler (kadınlar ve erkekler), çiftler, buz dansı ve senkronize olarak dört yarışma alanına ayrılmaktadır. Minikler (pre-novice), yıldızlar (novice), gençler (junior) ve büyükler (senior) olmak üzere dört seviyede yarışmalar yapılmaktadır. Tüm yarışmalarda (tekler, çiftler, buz dansı ve senkronize) her bir programı (kısa program, serbest program, kısa dans, serbest dans, zorunlu danslar) oluşturan öğelerin her biri için değerlendirme ölçeği (Scale of value) kullanılmaktadır. Yarışmalarda 12 tane hakem (son düzenlemelerle birlikte 9 hakem) yer almaktadır, bu hakemler Uluslararası Buz Pateni Federasyonu (International Skating Union) tarafından seçilmektedir. Tekler ile çiftlerde kısa ve serbest programlar yer alırken, buz dansında kısa dans, serbest dans ve zorunlu danslar bulunmaktadır (ISU, 2010).

Türkiye’de, genellenebilirlik kuramı ve çok değişkenlik kaynaklı Rasch modeli ile gerçek veriler kullanılarak yapılan çok az sayıda araştırma bulunmakta ve bu araştırmaların çoğunluğu eğitim alanındaki uygulamaları içermektedir. Spor alanında, genellenebilirlik kuramı ve çok değişkenlik kaynaklı Rasch modelinin birlikte kullanıldığı bir araştırmaya rastlanmamıştır. Bu açıdan bakıldığında spor alanında elde edilen verilerin kullanıldığı, tüm bu yaklaşımların birlikte yer aldığı ve gerçek verilerin kullanıldığı bir araştırmanın yapılmasına gereksinim olduğu söylenebilir.

Bunun yanında, performansın belirlenmesinde kullanılan hakem sayısında süreç içerisinde ekonomik nedenlerle yapılan değişikliklerin uygunluğunun değerlendirilmesini hedefleyen bir araştırmanın yapılmasının da önemli olduğu düşünülmektedir.

Bu arařtırmada, performansın ölçülmesinde birden fazla puanlayıcının puanlaması sonucu elde edilen puanlara genellenebilirlik kuramı ve çok deęişkenlik kaynaklı Rasch modelinin birlikte uygulanması ile hangisinin daha kullanılabilir sonuçlar verdiğinin saptanması, avantajlı olanının ve daha çok bilgi saęlayanının belirlenmesi, ayrıca, karar çalışmalarının geçerliğinin karşılaştırılması amacıyla ařağıdaki problem cümlesi ve alt problemlere cevap aranmaktadır.

1.6. PROBLEM CÜMLEĐİ

Birden fazla hakem tarafından puanlanan Dünya Buz Pateni Şampiyonasında tekler ve çiftler yarışmalarında elde edilen puanların genellenebilirlik kuramı ve çok deęişkenlik kaynaklı Rasch modeli ile kestirilen parametreleri nasıldır?

1.7. ALT PROBLEMLER

1. Buz pateni performansının ölçülmesinde, 2006–2011 Dünya Buz Pateni Şampiyonalarından elde edilen verilerle yapılan genellenebilirlik çalışmaları sonucunda kestirilen varyans bileşenleri ve toplam varyansı açıklama yüzdeleri nasıldır?
2. 2006–2008 ve 2009–2011 yılları arasında elde edilen verilerde orijinal hakem sayısı ile hakem sayısının artırılıp azaltılması ile yapılan karar çalışması sonucunda elde edilen G ve Phi katsayıları nasıl deęişmektedir?
3. 2006- 2008 yılları arasında buz pateni performansının ölçülmesinde 12 hakemin kullanıldığı yarışmalarda:
 - 3.1 Model veri uygunluęu nasıldır?
 - 3.2 Sporcuların yetenekleri ve uygunluk istatistikleri nasıldır?
 - 3.3 Hakemlerin katılık/cömertlik istatistikleri nasıldır?
4. 2009- 2011 yılları arasında buz pateni performansının ölçülmesinde 9 hakemin kullanıldığı yarışmalarda:
 - 4.1 Model veri uygunluęu nasıldır?

4.2 Sporcuların yetenekleri ve uygunluk istatistikleri nasıldır?

4.3 Hakemlerin katılık/cömertlik istatistikleri nasıldır?

5. Genellenebilirlik kuramı ve çok değişkenlik kaynaklı Rasch ölçme modeline göre birey (sporcu) ve hakem değişkenlik boyutlarında elde edilen parametrelerin benzerlik ve farklılıkları nelerdir?

1.8. SINIRLIKLAR

Araştırma, 2006–2011 Dünya Buz Pateni Şampiyonası tekler ve çiftler kategorileri serbest programda elde edilen program bileşenlerine ilişkin verilerle sınırlıdır. Buz pateninde iki farklı ölçekte değerlendirme yapılmaktadır. Teknik puanlama 7’li ölçekten (-3, +3) oluşmaktadır. Program bileşenleri ise 40 puanlı (0.25–10.00 arası, .25’lik artışlarla değerlendirme yapılmaktadır) ölçekten oluşmaktadır. Program bileşenlerine ait puanlar daha subjektif olduğundan, hakemler tarafından kolayca değiştirilebilmektedir (ISU, 2010). Bu nedenle bu araştırmada program bileşenleri dikkate alınmıştır.

Kısa program, buz dansı ve senkronize kategorileri araştırma kapsamında değildir.

1.9. ARAŞTIRMANIN AMACI ve ÖNEMİ

Bu araştırmada, genellenebilirlik kuramı ve çok değişkenlik kaynaklı Rasch modeli ile gerçek veriler kullanılarak kestirilen parametreler incelenerek iki yaklaşım karşılaştırılmıştır. Aynı veriler kullanılarak ÇDKRM’ne dayalı ölçme yöntemi ve genellenebilirlik kuramından kestirilen parametrelerin karşılaştırılarak gerçek verilerde hangi yaklaşımın avantajlı olduğu, daha çok bilgi sağladığı ve iyi işleyip işlemediği ile bu yaklaşımların ürettiği bilgilerin birbirleriyle tutarlılıkları araştırılmıştır. Genellenebilirlik kuramına göre performans puanlamada çaprazlanmış genellenebilirlik desenlerinin ve bu desenlerle yapılan karar çalışmalarının sonuçlarının karşılaştırılması ve karar çalışmalarının geçerliğinin test edilmesi amaçlanmaktadır. Bu araştırmayla, benzer ölçme durumları için ortaya atılmış olan genellenebilirlik kuramı ve ÇDKRM yaklaşımlarının buz pateni yarışmalarından elde edilen gerçek veriler kullanılarak

birbirleriyle ve kendi içlerindeki tutarlılıkları test edilerek literatüre katkı sağlanabileceği düşünülmektedir.

Sporcuların performanslarının değerlendirilmesinde, puanların ne kadar güvenilir olduğu merak edilmektedir. Bunun yanısıra buz pateni performansının ölçülmesinde etkili olan değişkenlik kaynağının ne olduğunu ve ölçme hatasını en aza indirmek için ölçmenin nasıl yapılması gerektiğinin bilinmesi de önemlidir. Aynı zamanda bu noktalara temas edilerek, spor alanında çalışanları bu konularda aydınlatmak çalışmanın bir diğer amacını oluşturmaktadır.

Genellenebilirlik kuramı ve ÇDKRM birçok alanda önemli bilgi sağlamasına rağmen, spor alanında yaygın olarak kullanılmamaktadır. Araştırma ayrıca genellenebilirlik kuramının ve ÇDKRM'nin spor bilimlerinde kullanımını ve genellenebilirlik kuramında değişkenlerin çapraz olarak tasarlanması sonucu oluşturulan desenlerin kullanılabilirliğini örnekleme amacı da taşımaktadır.

1.10. TANIMLAR VE KISALTMALAR

KTK: Klasik Test Kuramı

G- Kuramı: Genellenebilirlik Kuramı

G-Çalışması: Genellenebilirlik Çalışması

K-Çalışması: Karar Çalışması

ÇDKRM: Çok Değişkenlik Kaynaklı Rasch Modeli

ISU: Uluslararası Buz Pateni Birliği (International Skating Union)

G-Katsayısı: Genellenebilirlik Katsayısı

Phi Katsayısı: Güvenirlik Katsayısı

G: Ayırma Oranı

1.11. İLGİLİ ARAŞTIRMALAR

Bu bölümde, performansın ölçülmesinde genellenebilirlik kuramı ve çok değişkenlik kaynaklı Rasch ölçme modelinin kullanıldığı çalışmalar, yurtiçi ve yurtdışı araştırmalar

olmak üzere iki başlık altında özetlenmiştir. Buna ek olarak buz pateni değerlendirmelerini inceleyen araştırmalara da ayrı bir başlık altında yer verilmiştir.

1.11.1. Yurtiçinde Yapılan Çalışmalar

Atılgan (2004), müzik öğretmenliği seçme sınavında, birden çok görev için bireylerin gözlenmesi ve puanlanması durumunda genellenebilirlik kuramı ve ÇDKRM ile kestirilen istatistikleri karşılaştırmıştır. G-kuramı ve ÇDKRM yaklaşımlarıyla elde edilen sonuçların kısmen tutarlı sonuçlar verdiği, puanlayıcı ve görev değişkenlik kaynağı için her ikisinde tutarlı sonuçlar ürettiğini tespit etmiştir.

Atılgan (2005), “Genellenebilirlik Kuramı ve Puanlayıcılar Arası Güvenirlik için Örnek Bir Uygulama” adını verdiği çalışmasında, genellenebilirlik kuramını tanımlamış, kuramla ilgili temel kavramları açıklamış, klasik test kuramına göre üstün olan yönlerini vurgulamış ve uygulanmasını da hipotetik olarak örneklendirmiştir. Puanlayıcılar arası güvenirliliğin belirlenmesi gibi hata kaynağının birden çok olduğu ölçme durumlarında, ölçmenin psikometrik özelliklerinin belirlenmesinde, ölçme araçlarının geliştirilmesinde, en uygun madde ve puanlayıcı sayısının belirlenmesinde genellenebilirlik kuramının kullanılmasının daha uygun olabileceği sonucuna varmıştır.

Yelboğa (2007), 2005–2006 yılları arasında iş performansı ölçeği kullanarak KTK ve genellenebilirlik kuramından elde edilen güvenirlilik katsayılarını karşılaştırmıştır. İş performansı ölçeğinde KTK ve G-Kuramına göre elde edilen güvenirlilik katsayılarının uyumlu sonuçlar verdiğini bulmuştur.

Güler (2008), KTK, G-Kuramı ve Rasch Modeli üzerine yaptığı çalışmada matematik başarısının ölçülmesinden elde edilen sonuçları karşılaştırmıştır. Matematik başarısını ölçmek için kullanılan ölçme araçlarının güvenilir sonuçlar verdiği, puanlayıcıların birbiriyle uyumlu puanlamalar yaptığı belirlenmiştir. Güvenirliliğin belirlenmesinde en az iki kuramdan yararlanılması gerektiğini tespit etmiştir.

Atılgan (2008)'in arařtırmasında, 2003–2004 öğretim yılında İnönü Üniversitesi Eğitim Fakültesi Güzel Sanatlar Eğitimi Bölümü Müzik Öğretmenliği Programı özel yetenek seçme sınavlarının ikinci aşamasının güvenilirliği arařtırılmıştır. 249 adaya üç alt boyutta uygulanan 13 görev (deřifre; 4 görev, söyleme; 5 görev ve çalma; 4 görev) dört puanlayıcı tarafından puanlanmıştır. Sonuç olarak bazı bireylerin bir maddeden diđerine bađıl durumlarının farklılaşmasının az olduđu, az da olsa bazı puanlayıcıların bazı bireyleri diđerlerine göre daha katı ya da cömert puanladıkları ve puanlayıcıların bireyleri bir görevden diđerine kararlı puanladıkları görölmektedir. Alternatif K-Çalışmaları ile G ve Phi katsayılarının artırılmak istenmesi durumunda, puanlayıcı sayısının artırılmasından çok madde sayılarını artırmanın daha uygun olduđu sonucuna varmıştır.

Deliceođlu (2009) futbolcuların teknik yetilerinin tespit edilmesinde kullanılan Futbol Yetilerine İliřkin Dereceleme Ölçeđi (FYİDÖ)'nden elde edilen ölçümlerde (4 puanlayıcı); KTK ve G-kuramına dayalı olarak belirlenen güvenilirlik katsayılarının düzeylerini saptamış ve birbirleriyle karşılařtırmıştır. FYİDÖ'nün puanlamasından elde edilen güvenilirlik katsayıları iç ölçütlere göre incelendiđinde, G ve Cronbach Alfa katsayılarının beklenen deđerlerinden yüksek olduđu görölmektedir. Phi katsayısı ile Kendall W Güvenirlik katsayılarının beklenen deđerlerinden düşük olduđu görölmektedir. Potansiyel hata kaynaklarının fazla olduđu durumlarda G-Kuramı'nın, KTK'na alternatif oluşturduđunu söylemektedir.

Nalbantođlu (2009), tıp fakültesinde performans puanlamada öğrencilerin birden fazla puanlayıcı tarafından birlikte ve dönüşümlü olarak puanlanmasıyla oluşturulan desenlerde G ve K-çalışmaları sonuçlarını karşılařtırmıştır. Elde edilen sonuçlara göre öxgxp ve (ö:p)xg desenleri ile yapılan analizler sonucunda her iki desende deđişkenler için kesitirilen varyans deđerlerinin birbiriyle paralellik gösterdiđi bulunmuştur. Çok sayıda öğrencinin bulunduđu performans sınavlarında puanlayıcılar arası tutarlık sađlandığında, öğrencilerin puanlayıcıların hepsi tarafından tek tek puanlanması yerine puanlayıcıların belli sayıdaki öğrencileri dönüşümlü olarak puanlamasının zaman, iş gücü ve ekonomiklik açısından daha uygun olduđu sonucuna varmıştır.

Güler ve Çetin (2010), yaptıkları araştırmada 24 maddeden oluşan çoktan seçmeli bir sınavda farklı sayıda gözetmen kullanıldığı durumlarda, tek değişkenlik boyutunda öğrencilerden elde edilen puanların güvenilirlik katsayılarını klasik test kuramı ve genellenebilirlik kuramına göre hesaplamıştır. Tek değişkenlik kaynaklı çalışmada güvenilirlik değerlerinin yüksek olduğu gözlenmiştir. İki değişkenlik kaynağının birlikte ele alındığı çalışmada G ve Phi katsayıları, tek değişkenlik kaynaklı çalışmalardan elde edilen değerlerden daha büyük değerlere ulaşmıştır.

Güler ve Gelbal (2010) “Klasik Test Kuramı ve Çok Değişkenlik Kaynaklı Rasch Modeli Üzerine Bir Çalışma” adlı çalışmalarında matematik başarısında hakemler arası güvenilirliğin belirlenmesinde klasik test kuramına dayalı ölçme yöntemi ve çok değişkenlik kaynaklı Rasch modeli (ÇDKRM) yaklaşımlarını karşılaştırmıştır. Klasik test kuramına dayalı ölçme yöntemlerinden Cronbach alfa, Kendall uyum katsayısı, pearson korelasyon ve tek değişkenli varyans analizi kullanılmıştır. Elde edilen değerler puanlayıcılar arası uyumun olduğunu destekler niteliktedir. Facets analizi ile hesaplanan ayırma indeksinin oldukça yüksek bir değerde olduğu belirlenmiştir. Bu değere göre puanlayıcıların cömertlik/katılık açısından birbirlerinden farklı oldukları sonucuna varılmıştır. Bununla birlikte, uyum istatistiklerinin işaret ettiği şekliyle; bir bütün olarak puanlayıcıların tutarlı puanlama yapmış oldukları belirlenmiştir. ÇDKRM, tek bir çalışma ile her bir değişkenlik kaynağına ilişkin farklı güvenilirliklerin hesaplanmasına imkân verdiğinden, her bir değişkenlik kaynağı için ayrıntılı bilgi sağlamaktadır. Tüm değişkenlik kaynaklarının tek bir ölçekte birlikte değerlendirilmesi, her bir değişkenlik kaynağı içindeki uyumsuzlukların belirlenmesi ve böylece hangi madde ya da puanlayıcıdan kaynaklı bir sorun olduğunun açıkça görülmesinin istendiği durumlarda ÇDKRM'nin kullanımının uygun olduğunu belirtmişlerdir.

Öztürk (2011), KTK ve G-Kuramı üzerine yaptığı çalışmada voleybolcuların teknik becerilerinin tespit edilmesi için kullanılan voleybol becerilerine ilişkin dereceleme ölçeğinden elde edilen ölçmeleri karşılaştırmıştır. KTK (Cronbach alfa, Kendall uyum katsayısı) ve Genellenebilirlik Kuramına (genellenebilirlik ve güvenilirlik katsayıları) dayalı olarak belirlenen güvenilirlik katsayıları hesaplanmış ve birbirleriyle karşılaştırmıştır. Elde edilen sonuçlara göre, aynı ölçme durumu için KTK ve G-

Kuramı'ndan elde edilen güvenilirlik katsayılarının birbirleriyle uyumlu sonuçlar ürettiğini bulmuştur.

Yurtiçinde yapılan çalışmalar incelendiğinde, klasik test kuramı, genellenebilirlik kuramı ve ÇDKRM'nin tutarlı sonuçlar verdikleri bulunmuştur. Genellenebilirlik kuramı ve ÇDKRM'nin potansiyel hata kaynaklarının fazla olduğu durumlarda kullanımının daha uygun olduğu belirlenmiştir.

1.11.2. Yurtdışında Yapılan Çalışmalar

MacMillan (1995) klasik test kuramı, genellenebilirlik kuramı ve ÇDKRM'ni büyük ölçekli performans değerlendirme durumlarında puanlayıcıların katılığı, tutarlılığı ve uyumu açısından karşılaştırdığı çalışmasında 3 ayrı göreve katılan 4930 katılımcı, 70 puanlayıcı tarafından puanlanmıştır. Her bir katılımcı 70 kişilik puanlayıcı havuzundan seçilen 3 puanlayıcı tarafından puanlanmıştır. Bu araştırma sonucunda, araştırmacı üç yaklaşımın her durumda aynı sonuçlar üretmediğini ve hangi yaklaşımın veya yaklaşımların kullanılacağına amaca göre seçilmesinin önemli olduğu vurgulamıştır.

Lunz ve Schumacker (1997) performans sınavından elde edilen verileri yorumlamak ve sonuçları karşılaştırmak için dört tane yöntem kullanmışlardır. Kullandıkları yöntemler: (a) geleneksel betimleyici istatistikler, (b) puanlayıcılar arası korelasyonlar, (c) genellenebilirlik kuramı ve (d) ÇDKRM. Karşılaştırmalar sonucunda her yöntemde benzer değişkenlik kaynakları belirlenmiştir ancak ÇDKRM doğrusal ölçekte puanlar üretebilmekte ve adayların girdikleri sınav farklılıklarını da dikkate almaktadır. Bu tür performans sınavlarında genellikle 4 tane değişkenlik kaynağı bulunmaktadır. Bunlar (a) bireyin yeteneği, (b) görevler veya puanlama boyutları, (c) maddeler veya başlıklar ve (d) puanlayıcılar. Lunz ve Schumacker (1997) beş değişkenlik kaynaklı yuvalanmış desende ÇDKRM kullanmışlardır. Bu modelin içerdiği değişkenlik kaynakları: bireyler, başlıklar, görevler, puanlayıcılar ve seanslardır. Bu desende, her bir değişkenlik kaynağının bireyin yetenek kestirimine etkisi değerlendirilmektedir (Lunz ve Schumacker, 1997). Araştırma sonucunda, bireylerin puanlandığı seansların,

puanlayıcıların (altı tane), başlıkların ve görev güçlüğü'nün bireyin aldığı puanı etkilediği ortaya koyulmuştur.

Stahl, Shumway, Bergstrom ve Fisher (1997) tarafından yapılan çalışmada, mesleki terapistler tarafından iyileştirici terapi programının danışanlar üzerindeki etkililiğini belirlemede kullanılan performans değerlendirme aracından elde edilen veriler kullanılmıştır. Bilgisayar ortamındaki bu sistemi ve ÇDKRM'ni kullanmanın; (a) standartlaştırılmış durumlarla puanlayıcıları kalibre etmeyi, (b) puanlayıcı eğitimi sırasında yeni puanlayıcıların kalibrasyonunu yapmayı, (c) değerlendirme sürecindeki her bir değişkenlik kaynağının bağımsız kestirimine olanak sağlayarak bireyin yeteneğini değerlendirmeye odaklanmayı, (d) değerlendirmeyi takiben bireyin yeteneğini anında tahmin etmeyi olanaklı kıldığını belirlemiştirlerdir.

Lynch ve McNamara (1998) göçmenlerin konuşma becerilerine ilişkin performans testi geliştirmede ÇDKRM ve genellenebilirlik kuramını kullanmıştır. ESL konuşma becerilerinden elde edilen verileri kullanarak bu iki modeli karşılaştırmışlardır. Çalışma sonucunda, genellenebilirlik kuramının test deseni, testin uzunluğu ve gerekli hakem sayısı ile ilgili genel kararlar verme gibi genel ve grup düzeyinde bilgi sağladığını ortaya koymuşlardır. Mikroskop benzetmesi kullanarak karşılaştırma yapmışlardır. Buna göre, ÇDKRM'nin yüksek düzeyde büyütebildiğini (mikroskop gibi düşünürsek daha detaylı inceleme) ve araştırmacıya tüm kusurları açıklama olanağı sunduğunu, buna karşın, G-kuramının büyütme düzeyinin düşük olduğunu söylemişlerdir. Bu çalışmada, ÇDKRM birçok birey- puanlayıcı ve puanlayıcı-madde etkileşiminin yanlı olduğunu gösterirken, G-çalışması ise bu yanlılıkların toplam düzeyde iptal edildiğini ortaya koymuştur. Grup düzeyindeki istatistikleri ve K-çalışmaları ile genellenebilirlik kuramı test deseni ile ilgili kararlar vermede yararlıyken, ÇDKRM'nin ise sağladığı bilgi ile puanlayıcılar ve maddelerde düzeltmeler yapmaya olanak sağladığı ortaya koyulmuştur.

MacMillan (2000) KTK, ÇDKRM ve genellenebilirlik kuramını karşılaştırmış ve her üç kuramda da puanlayıcı değişkenliğinin etkili bir şekilde belirlenebildiğini bulmuştur. ÇDKRM, genellenebilirlik kuramına göre daha fazla değişimi açıklarken,

genellenebilirlik kuramının sadece deęişkenlik kaynađını deęil, deęişkenlik kaynakları arasındaki etkileşimleri de dikkate aldığını bulmuştur. ÇDKRM'nin ise etkileşimin olmadığını varsaydığını ve her bir deęişkenlik kaynađını bağımsız olarak ele aldığını ileri sürmektedir.

Kozaki (2004), Japonca'dan İngilizce'ye tıbbi tercüme yapmanın sertifikalandırılması amacıyla kullanılan performans deęerlendirmelerinde çoklu standartlar belirlemek için genellenebilirlik kuramı ve ÇDKRM'ni karşılaştırmıştır. Genellenebilirlik kuramı bireyin yetenek kestiriminde; deęişkenlik kaynaklarına, deęişkenlik kaynakları arasındaki etkileşimlere ve ölçme nesnesine (amacına) yüklenebilecek görelî etkileri belirlediğini, ÇDKRM'nin ise deęişkenlik kaynaklarının belirli bir ögesine ilişkin bilgi sunduđunu ortaya koymuştur. Örne; belirli bir deęişkenlik kaynađı ve etkileşimler tarafından oluşturulan deęişkenlikten hangi hakemin, maddenin veya etkileşimlerin daha fazla sorumlu olduđuna ilişkin bilgiler vermektedir. Puanlayıcı özellikleri ile ilgilenen araştırmalarda ÇDKRM, G-Kuramından daha uygun bir teknik olduđu söylemektedirler. Çünkü ana etkinin ve etkileşimlerin etkisinin ötesine giderek, bireysel düzeyde puanlayıcı etkilerinin araştırılmasına olanak sağladığı ileri sürülmüştür.

Smith ve Kulikowich (2004) karmaşık problem çözme becerilerinin deęerlendirilmesinden elde edilen puanları kullanarak genellenebilirlik kuramını ve ÇDKRM'ni karşılaştırmışlardır. Deęişkenlik kaynakları arasındaki deęişimin görelî büyüklüđünün karşılaştırılabilir olduđunu ancak varyans kaynaklarını ele alma şekillerinin farklı olduđunu bulmuşlardır.

Sudweeks, Reeve ve Bradshaw (2005), üniversite ikinci sınıf öğrencilerinin yazılarının deęerlendirilmesinde G-kuramı ve ÇDKRM'ni karşılaştırmışlardır. G-kuramı analiz edilen veriyi temel etki ve etkileşimler ile açıklayan kapsamlı bir yaklaşım olarak tanımlarken, ÇDKRM'ni analizdeki deęişkenlik kaynaklarına tek tek odaklanan dar bir yaklaşım olarak tanımlamaktadırlar. G-kuramının ve ÇDKRM'nin kendilerine özgü güçlü yanları olduđunu belirtmişlerdir. İki tekniğin odak noktalarının farklı olduđunu ve hangisinin uygun olduđuna ilişkin verilecek kararların araştırma bağlamına bađlı

olduğunu savunmaktadırlar. Her iki analizden elde edilecek sonuçların birbirlerini tamamlamada kullanılabilecekleri sonucuna varmışlardır.

Alharby (2006) puanlamada iki farklı yaklaşımı (bütüncül ve analitik) karşılaştırdığı çalışmasında, ölçümlerin güvenilirliğini de iki ayrı teknikle karşılaştırmıştır. Bunlar; genellenebilirlik kuramı ve ÇDKRM'dir. Çalışmada, iki puanlama yaklaşımı ile güvenilirlik yöntemleri arasındaki etkileşimleri de ele almıştır. Genellenebilirlik çalışması sonucunda analitik puanlama yönteminin bütüncül puanlamaya göre daha yüksek G-katsayısına sahip olduğunu bulurken. ÇDKRM uyguladığında bütüncül yaklaşımın, analitik yaklaşıma göre daha iyi uyum gösterdiğini bulmuştur. Her iki modelin de birbirine karşı avantajları ve dezavantajları olduğunu ve hangi modelin kullanılacağına ilişkin kararın araştırmacının amacına bağlı olduğunu, daha ayrıntılı ve kapsamlı bir sonuç elde etmek için her iki modelin de kullanılabileceğini söylemektedir.

Kim ve Wilson (2009) öğrencilerin yazma becerilerinin değerlendirilmesi sonucunda elde edilen veriler ile yaptıkları çalışmada genellenebilirlik kuramı ve ÇDKRM karşılaştırmıştır. Her iki kuramın da amaca bağlı olarak önemli katkılar koyduğunu söylemektedirler. Her iki ölçme yönteminin de bireyler, maddeler ve puanlayıcılar hakkında bilgi verdiğini, araştırmacıların neyi çalışmak istediklerini net bir şekilde belirlemeleri gerektiğinin önemli olduğu belirtilmiştir. Eğer araştırmacı, grupların nasıl hareket ettiği, çalıştığı (behaving) ile ilgileniyorsa G-kuramının daha yararlı olacağını; eğer araştırmacı bireysel performans ile ilgileniyorsa ÇDKRM veya bir diğer MTK yaklaşımının kullanılmasının daha iyi olacağını savunmaktadırlar.

Sebok (2010) Northern British Columbia Üniversitesi'nin, Eğitimde Danışmanlık Yüksek Lisans Programına öğrenci seçme sürecini ÇDKRM ve genellenebilirlik kuramı kullanarak değerlendirmiştir. Araştırmacı iki yöntemin de puanlayıcılar arası uyumu değerlendirmede uygun olduğunu ancak iki yöntemin de avantajları ve dezavantajları olduğunu söylemektedir. ÇDKRM'nin küçük örneklerde ve eksik verilerle kullanabilmesinin avantaj sağladığını söylerken, en uygun örneklem büyüklüğü ve uyum istatistikleri gibi konularda somut kurallarının eksik olmasını dezavantajları arasında saymaktadır. Genellenebilirlik kuramının ise, her bir değişkenlik kaynağının

ana etkisi ve olası tüm etkileşimlerine ilişkin varyans bileşenlerini verdiği, görelî veya mutlak kararlar için ayrı katsayılar hesapladığını ve çeşitli desenler için karar çalışmalarına olanak tanıdığını ancak eksik verilerle işlem yapamadığını belirtmiştir.

Yang (2010) yaptığı araştırmada zaman içerisinde Sözel İngilizce Yeterlik Testinde (Kuzey Amerika Üniversitesinde Uluslararası Öğretim Asistanlarını değerlendirmede kullanılan bir test) bireylerin puanlarında puanlayıcı katılımının etkisini araştırmıştır. Ağustos 2007’de 9 ve 2008’de 10 puanlayıcının değerlendirme yaptığı sınav sonuçlarından elde edilen verileri Facets programını kullanarak ÇDKRM ile analiz etmiştir. Araştırma sonucunda puanlayıcıların farklı düzeylerde katılım gösterdikleri bulunmuştur. Ancak test puanlarında hakem katılımının etkisi oldukça küçüktür. İki uygulamadaki bireylerin yaklaşık %4’ü beklenen ortalamadan daha yüksek gözlenen ortalamaya sahiptirler. Puanlayıcıların çoğu ölçeği tutarlı bir biçimde kullanmışlardır. Bir puanlayıcı, en üst kontrol sınırı olan 1.2’den biraz daha yüksek bir iç uyum istatistiği ile tutarsızlık göstermiştir. Hakemlerin puanlamaya ilişkin katılımları iki dönemde aynı kalmasa da kabul edilebilir aralıklarda kalmıştır. İki puanlayıcı, Rasch modelinin beklediğinden daha fazla sapma göstermiştir. Yeni puanlayıcılar katılım açısından, tecrübeli puanlayıcılardan farklılık göstermemişlerdir. Puanlama kategorileri arasında farklılıklar biraz fazla bulunmuştur, bunun için de ölçeğin gözden geçirilmesi gerektiğini önermektedir. Bunların yanında, bu araştırma Facets’in puanlayıcı performansını değerlendirmede kullanışlı bir araç olduğunu ve programdan elde edilen sonuçların, puanlayıcıların doğru puanlamasını sağlamak amacı ile puanlayıcı eğitimlerini takiben kullanılabilceğini iddia etmiştir.

Ure (2011) dnyayıcı öğretime ilişkin performansın değerlendirilmesinde, iki farklı puanlama koşulunun (kontrollü ve kontrollü olmayan) puanlayıcı davranışını ve performans değerlendirmenin güvenilirliğini nasıl etkilediğini araştırdığı çalışmasında ÇDKRM ve genellenebilirlik kuramını güvenilirlik belirleme performansları açısından karşılaştırmıştır. Bu çalışma sonucunda araştırmacı, her iki modelin de katkılarının bulunduğunu ve iki modelin de kullanılmasının önemli olduğunu söylemektedir. Her iki modelden elde edilen güvenilirlik kestirimlerinin birbirinden istatistiksel olarak farklı olmadığını ancak genel bir bilgi edinmek amacı ile öncelikle g- kuramının

uygulanmasını ve onun ardına da ÇDKRM kullanılarak ayrıntıların incelenmesini önermektedir. Her iki modelin de farklı avantajlarından yararlanmanın önemi üzerinde durmuştur.

Yurtdışında yapılan çalışmalar incelendiğinde, genellenebilirlik kuramının genel ve grup düzeyinde bilgi sağladığı, ÇDKRM'nin ise her bir değişkenlik kaynağına ilişkin bireysel düzeyde ayrıntılı bilgi sağladığı bulunmuştur. Genellenebilirlik kuramının, değişkenlik kaynakları arasındaki etkileşimleri de dikkate aldığını ve karar çalışmaları ile gelecekteki çalışmalar için değişkenliği düzeltmeye yönelik olduğunu ortaya koymuşlardır. Çalışmalar, hangi tekniğin kullanılacağına amaca göre seçilmesinin önemi üzerinde durmuşlardır. Bu tekniklerin birbirini tamamladıklarını ileri sürmüşlerdir.

1.11.3. Buz Pateni ile İlgili Araştırmalar

Diğer birçok yarışma sporu gibi, buz pateninde puanlama, değerlendirmelerinde yansız olması beklenen, çok iyi eğitilmiş hakemler tarafından yapılmaktadır. Ancak yanlılık veya adil olmayan puanlamanın çeşitli nedenlerden dolayı ortaya çıktığına inanılmaktadır. Bunun nedenleri zevk, politik veya milliyetçi yönelimler, coğrafi bağlantı, yarışma sırası, kişisel tercih, sporcunun şöhreti ve/ya yarışma düzeyi olduğu iddia edilmektedir (Bassett ve Persky, 1994; Findlay ve Ste-Marie, 2004; Seltzer ve Glass, 1991; Wu ve Yang, 2004). Buz pateni yarışmalarında, belirli bir grup sporcu birden çok hakem tarafından puanlanmakta ve hakemler arası tutarlılığın derecesi önem kazanmaktadır (Saeki ve Fan, 1999). Sporcular, gerekli öğelerden oluşan kısa ve uzun programlarda performans göstermekte ve performanslarının hem teknik hem de sunum özellikleri birden fazla hakem tarafından puanlanmaktadır, bu da potansiyel bir tutarsızlığı getirebilmektedir.

Buz pateni yarışmalarındaki puanlamada da kullanıldığı gibi toplu kararlar verme, puanları bir araya getirme problemlerini de ortaya çıkarmaktadır. Tartışmalar iki uç arasında gidip gelmektedir. Bir yandan ölçümlerin objektif olduğu savunulurken bir yandan da ölçümlerin öznel olduğu ve yanlılığa açık olduğu savunulmaktadır. Spordaki

puanlamaların öznel doğası yıllardır spor haberlerinin başlıklarını oluşturmaktadır. Milliyetçi, politik ve coğrafi yakınlıkların hakemlerin kararlarını etkilediği çok uzun zamandan beridir şüphe uyandırmaktadır. Milletler arasındaki savaşlar, örneğin; dünya savaşları ve soğuk savaş, doğulu ve batılı ülkelerden gelen hakemlerin puanlamadaki sistematik yanlılığına yönelik soruları akıllara getirmektedir. 1967 ile 1971 yılları arasındaki Dünya Buz Pateni Şampiyonaları yanlılık açısından incelenmiştir. Ancak elde edilen bulgular jeopolitik yanlılık hipotezini desteklememiştir (Ball, 1973). 1968'den 1988'e kadar olan Olimpiyat oyunlarındaki veriler incelendiğinde ise hakemlerin “milliyetçi bağlardan veya politik bağlantılardan” etkilendikleri kanıtlanmıştır (Seltzer ve Glass, 1991). Daha yakın geçmişe bakacak olursak, 2002 Tuz Gölü Kış Olimpiyatlarında Fransız hakem, federasyon başkanının politik baskısı sonucunda yanlı oy kullanmak ile suçlanmıştır.

Buz pateni yarışmalarında puanlama güvenilirliğini araştıran ampirik araştırmalarda (Ball, 1973; Seltzer ve Glass, 1991; Weekley ve Gier, 1989; Whissell, Lyons, Wilkinson ve Whissell, 1993) kullanılan desenler ve yapılan analizler puanlamanın karmaşık doğasını yani çok değişkenli doğasını dikkate almamıştır. Örneğin; hakemler arası güvenilirlik kestirimleri KTK'na dayalı olarak yapılmıştır. KTK çerçevesinde, puanlayıcılar arası güvenilirlik kestirimi de genellikle iki hakem ile sınırlandırılmıştır. Ancak buz pateninde puanlamada daha fazla hakem kullanılmaktadır. Daha sonraki çalışmalar ise güvenilirlik kestirimlerinden Kendall'ın uyum katsayısını kullanmışlardır (ör; Ball, 1973; Seltzer ve Glass, 1991). Kendall'ın uyum katsayısı sadece sıralamanın güvenilirliğini açıklamakta ve orijinal ham puanların güvenilirliğini açıklamamaktadır. Sıralamaya dayanan güvenilirlik kestirimlerinin, orijinal puanlara dayananlara göre daha az doğru veya kesin olduğuna inanılmaktadır. KTK ve Kendall'ın uyum katsayısı ile ilgili çalışmalara cevap olarak Saeki ve Fan (1999) 1998 Nagano Kış Olimpiyatları buz pateni yarışma sonuçlarına ilişkin puanlama güvenilirliğini genellenebilirlik kuramı çerçevesinde değerlendirmiştir.

Wanderer (1987) ise yarışma sırasına dayanan yanlılığa kanıt sağlamıştır. Bölgesel ve ulusal buz pateni şampiyonalarına ilişkin gözlemleri sonucunda, önceki yarışmadaki

sonuçların yayınlanmasından sonra, hakemlerin daha sonraki yarışmalarda verdikleri puanları paneldeki tüm hakemlerin ortalamasına doğru geri çektiğini bulmuştur.

Weekly ve Gier (1989) yaptıkları çalışmada 1984 Olimpiyatlarında buz pateni (erkekler ve kadınlar) yarışmalarının sonuçlarını kullanmışlardır. Araştırmada, hakemler arası güvenilirliği oldukça yüksek bulmuşlardır (.93-.97). Bu kestirimler, iki hakeme göre yapıldığı zaman güvenilirlik katsayıları düşüş göstermesine (.81) rağmen yine de görece olarak yüksek bulunmuştur. Ayrıca bu çalışmada, her iki cinsiyette de hakemler sporcuları sıralarken, ilk sıralamalarda yer alan sporcular konusunda 3. ile 23. sporcuların sıralanmasına göre daha tutarlı görüş bildirmişlerdir.

Seltzer ve Glass (1991), 1968 ile 1988 yılları arasında yapılan kış olimpiyatlarındaki buz pateni yarışmalarından elde edilen verilerde, sporcuların değerlendirilmesinde politik etkilerin yerini belirlemeyi amaçlamışlardır. Milliyetçiliğin, buz pateni gibi subjektif değerlendirmelere dayanan sportlardaki etkisinin, genel performans standartlarında uyuşmazlıklar ve hakemlerin, kendi ülkesinden olan sporcuları kayırma eğilimi şeklinde görüldüğünü belirtmişlerdir. Buz pateni ve buz dansı (417 sporcu ve çift) yarışmalarından elde edilen veriler incelendiğinde, hakemlerin kendi ülkelerinden olan sporculara anlamlı bir şekilde daha yüksek puanlar verdiğini bulmuşlardır. Buna ek olarak, elde edilen sonuçlar milliyetçi politikaların veya soğuk savaşların puanlamaya rehberlik ettiğini bulmuşlardır. Örneğin, Federal Almanya Cumhuriyeti ile Alman Demokratik Cumhuriyeti'nden gelen hakemler Alman sporculara ortalamanın üzerinde (1.13 ve 1.20) puanlar verirken, Sovyetler birliğinden gelen hakem, Amerikalı sporcuya ortalamanın altında (-1.25) puan vermiştir.

Whissell, Lyons, Wilkinson ve Whissell (1993), 1984 ve 1988 Olimpiyatlarında yer alan buz pateni yarışmalarının çeşitli bölümlerine ilişkin hakem puanlarını yanlılık açısından dört yöntemle (ortalamalara dayalı, maksimum frekans yöntemi, minimum frekans yöntemi ve sıra sapma yöntemi) incelemişlerdir. Bunun sonucunda "ulusal yanlılık" yapıldığı sonucuna varmışlardır. Hakemler, kendi ülkelerinden olan sporcuları yanlı puanlamaktadır. Hakemler, kendi ülkelerinden olan sporculara, diğer hakemlerin

verdikleri puanların ortalamasından daha yüksek puanlar vermişlerdir. Sporcuların elde ettikleri toplam değerlendirme puanının üzerinde puan verdikleri bulunmuştur.

Bassett ve Persky (1994) ise yarışma düzeyinin hakemin tutarlılığını etkilediği yönünde bir hipotez ortaya atmışlardır. Bu araştırmacılara göre, olimpiyat oyunları ve dünya şampiyonası gibi daha prestijli yarışmalardaki puanlamalar hakemlerin zevkinden ve ölçme hatalarından etkilenmektedir. Ulusal yarışmalardaki ölçümler ise deneyim, bağlılık ve hakemlerin profesyonelliği gibi problemlerden dolayı şüpheli olabileceği ileri sürmüşlerdir.

Campbell ve Galbraith (1996) hakemlerin bireysel olarak verdikleri puanlar üzerinde durmuştur. Her bir hakemin her sporcuya verdikleri puanların yanı sıra, dokuz hakemin verdikleri puanlara ilişkin ortalama ve medyanı da hesaplamışlar, buna ek olarak, sporcular ile hakemlerin ülkeleri ve sporcuların nihai sırasını da vermişlerdir. Ortalamalar ve medyanlar belirli bir sporcunun performansına ilişkin değerlendirmelerin sonuçlarını içerdiğinden dolayı, sporcular arası farklılıklar veya farklı ülkelerden gelen sporcular arasındaki sistematik farklılıklar dolaylı olarak kontrol edilmiştir. Sonuçlar incelendiğinde, hakemlerin kendi ülkesinden olan sporcuya olası en yüksek puanı verdiği (5.6 yerine 5.7) bulunmuştur. Yabancı bir ülkeden olan sporcuya da, hakemlerin rastgele bir biçimde yaklaşık olarak eşit oranda en yüksek veya en düşük puanı verdiklerini ortaya koydular. Böyle bir değerlendirmenin 0.05'lik bir yanlılık gösterdiğini, çok küçük de olsa aynı ülkeden olan hakem tarafından avantaj kazanıldığını söylemektedirler. Bu sonuçları yorumlarken birkaç noktaya dikkat çekmişlerdir. Öncelikle, yarışmalarda, birçok sporcunun kendi ülkesinden bir tane hakem (9 kişilik hakem heyeti arasında) bulunduğunun ve Kış Olimpiyatlarında birçok yarışmacının birkaç ülkenin oluşturduğu çok küçük ülkeler kümesinden geldiğinin altını çizmişlerdir. Bu kapsamda, farklı ülkelerden gelen hakemlerin eşit olarak yanlılık gösterdiğini ve bu avantajın da birçok sporcu için etkisiz hale geldiğini belirtmişlerdir. İkinci olarak ise, Bassett ve Persky (1994)'nin çalışmalarında elde edildiği gibi, son sıralama sisteminin bu tür yanlı bir sıralamaya karşı oldukça güçlü ve dirençli olduğunu, yani bu yanlılıktan etkilenmediğini belirtmişlerdir. Son olarak, bu çalışmada ortaya çıkan bir yanlılığın nedeni de hakemlerin ve sporcuların ulusal zevklerinin ortak

olmasının olabileceğini iddia etmişlerdir. Böylelikle “A” ülkesinden gelen bir sporcunun serisinde bulunan hareketlere ilişkin performansının “A” ülkesinden gelen hakemin hoşuna gidebileceği ve böyle bir yorumlamanın, hakemin kendi ülkesindeki sporcuyla kayırdığı anlamına gelmediğini söylemektedirler.

Looney (1997), 1994 yılında yapılan Buz Pateni Dünya Şampiyonası kadınlar finalinde elde edilen verilerde kısmi puanlama modelini (partial-credit rating scale) kullanmıştır. Analizler sonucunda elde edilen istatistiklere göre teknik programın, serbest programa göre daha az yordanabilir olduğunu belirtilmiş ve teknik programın değerlendirmesinin daha zor olmasının buna neden olduğunu ileri sürmüştür. 459 tane sıralamadan sadece 5 tanesinin uyumsuzluk gösterdiği bulunmuştur. Bunlar teknik programda ortaya çıkmış ve bir hakemin bu sporcuları beklenenden daha yüksek puanladığı belirlenmiştir. Birçok hakem oldukça yüksek iç tutarlık gösterirken, iki hakemin (İngiliz ve Kanadalı) puanlamalarındaki değişkenlik diğerlerine göre daha fazla bulunmuştur. Hakemler iki sporcuda anlaşmazlık yaşamışlardır. İngiliz ve Amerikan hakemler Çinli sporcuya teknik programda beklenenden daha yüksek puanlar verirken, Ukraynalı hakem ise Çek sporcuya her iki görevde de (teknik ve serbest) yüksek puanlar vermiştir. Sonuç olarak, bu iki sporcunun puanları oldukça yüksektir. Ancak bunların uygulamadaki sonuçları, sporcuların madalya için yarışan sporcular olmamasından dolayı önemli bulunmamıştır. Birçok sporcunun puanlamaları modele uyum gösterdiğinden, sporcuların yetenek ölçümlerinin, kayma becerilerinin anlamlı bir belirleyicisi olduğu belirtilmiştir.

Saeki ve Fan (1999) 1998 Nagano Kış Olimpiyatları buz pateni yarışma sonuçlarına ilişkin puanlama güvenilirliğini genellenebilirlik kuramı çerçevesinde değerlendirmiştir. Bunun sonucunda Saeki ve Fan (1999) birden çok hakemden elde edilen puanların genellenebilirlik katsayılarını oldukça yüksek (0.98 ile 0.99) bulmuştur. Sonuç olarak, 1998 Kış Olimpiyatlarında buz pateninde kullanılan puanlama sisteminin oldukça güvenilir olduğunu söylemişlerdir. Bu çalışma spor ortamındaki puanlamalarda genellenebilirlik kuramının kullanılmasında bir aşama olmuştur, ancak bu çalışma tek değişkenlik kaynaklı desenle (sporcu- hakem) sınırlandırılmış ve ayrı ayrı analizler yapılmıştır. Bu çalışmada, serbest ve kısa program ve puan boyutları (teknik ve sunum puanları) dikkate alınmamıştır.

Hellsten, Henderson ve Klinger (2000) aynı veriyi kullanarak daha karmaşık ve çok değişkenlik kaynaklı bir analiz yapmışlardır (sporcu-program-kategori-puanlayıcı), Hellsten ve diğ. (2000) 1998 Nagano Kış Olimpiyatlarındaki buz pateni yarışmasının hem puanlama değişkenliğini hem de güvenilirliğini açıklamışlardır. Bu çalışma sonucunda, sporcu varyansının %67 ile %91 arasında değiştiğini ve en yüksek değerin zorunlu dansta olduğu bulunmuştur. Genellenebilirlik katsayıları ise 0.95 ile 0.99 arasında bulunmuştur. Dans yarışmasında, özellikle de hakemlere göre puan varyansının en yüksek olduğu zorunlu dansta hakemin görelî önemi ve etkileşimler problemlî bulunmuştur.

Zitzewitz (2002, 2006) Olimpik Kış Sporları üzerine yaptığı araştırmasında kayakla atlama ve buz patenini ayrı ayrı incelemiştir. Bu spor dallarında da hakemlerin, kendileri ile aynı ülkeden gelen sporculara yönelik öznel kararlarında yanlılık gösterdiği bulunmuştur. Öznel kararların verildiği durumlarda genellikle yanlılık problemi ile karşılaşıldığı ve buna genel bir çözüm olarak birden fazla değerlendiriciyi karar verme sürecine katılması gerektiğini söylemektedirler. Ancak bu durumda da bu sürecin nasıl tasarlanacağına ilişkin problemler yaşandığını belirtmişlerdir.

Findlay ve Ste-Marie (2004) sporcunun ününün puanlamaya etkisini araştırdıkları çalışmalarında, Kanadalı 12 hakemin (6 Ontorio, 6 Quebec hakemi), buz pateni kısa programda yarışan 14 kadın sporcuyu (Quebec hakemlerinin iyi bildiği 7 kadın sporcu ile Ontorio hakemlerinin iyi bildiği 7 kadın sporcu) puanlamasına dayanan araştırmalarında videolardan yararlanmışlardır. Araştırma sonucunda daha iyi bilinen sporculara yönelik kayırmacılık bulmuşlardır. Teknik puanlarda bilinen sporcuların anlamlı bir şekilde daha yüksek puanlar aldığı bulunurken, artistik puanlarda fark bulunmamıştır.

Huang ve Foote (2011) yaptıkları çalışmada buz pateni yarışmalarında yapılan puanlamanın güvenilirliğini değerlendirmek amacı ile 2004 Dünya Buz Pateni Şampiyonasında yer alan dört kategoriye (tekler- kadınlar ve erkekler, çiftler, buz dansı) ilişkin verileri genellenebilirlik kuramını kullanarak analiz etmişleridir. Araştırma sonuçlarına göre, tekler kategorisinde hakem değişkenlik kaynağının toplam varyansın

yaklaşık %10'unu açıkladığını bulmuşlardır. Bunu da puanların güvenilirliğinde hakemlerin önemli olduğu şeklinde yorumlamışlardır. Ayrıca genellenebilirlik katsayısı tek kadınlar ve tek erkeklerde 0.93 olarak bulunmuştur. Çiftler ve buz dansı kategorilerinde ise elde edilen genellenebilirlik katsayıları sırası ile 0.99 ve 0.98 olarak bulunmuş ve bu iki kategorideki güvenilirliğin oldukça yüksek olduğunu ortaya koymuşlardır.

Looney (2012), 2010 Kış Olimpiyatlarında erkekler serbest program yarışmasından elde edilen verilerle yaptığı araştırmasında, beş program bileşenini ÇDKRM ile analiz etmiştir. Araştırmada; hakemlerin puanlarının model uyumu, hakemlerin “geçiş” bileşenini puanlamada güçlük yaşayıp yaşamadıkları, Olimpiyatlar öncesinde Plushenko'nun yaptığı basın açıklaması sonucunda bu sporcuyu puanlamada güçlük yaşayıp yaşamadıkları ve ISU koridor yöntemi ile Rasch analizlerinin aynı puanlayıcı hatalarını ortaya koyup koymadığı problemleri ele alınmıştır. Sonuç olarak, beş program bileşenindeki puanların model veri uyumunu sağladığını bulmuşlardır. Geçiş bileşeni hakemler tarafından çok yüksek puanlar verilmeyen ve aynı zamanda hakemlerin en çok sorun yaşadıkları bileşen olarak bulunmuştur. En önemli sonuçlardan bir tanesi de, Plushenko'ya verilen puanlara ilişkin Rasch ölçme sonuçları veya hakemlerin ham puanlarının ortalaması, Plushenko ile madalya için yarışan diğer sporcuların kayma yeteneğine ilişkin aynı yapıları yansıtmadığı ortaya koyulmuştur. Üç hakemin, Plushenko'yu geçiş bileşeninde beklenenden 1.5-2 puan daha düşük değerlendirdiği bulunmuştur. Bunun yanında, hakemlerden bir tanesinin geçiş bileşeninde daha düşük değerlendirmesine rağmen, performans ve yorumlama bileşenlerinde daha yüksek değerlendirdiği belirlenmiştir. Araştırma sonucunda, ISU'nun koridor yöntemi ile Rasch modelinin bir arada kullanımının puanlayıcı performansına ilişkin daha doğru bir çerçeve sağlayacağı sonucuna varılmıştır.

Buz pateni hakem değerlendirmeleri ile ilgili yapılan araştırmalar incelendiğinde, puanlamalarda bir miktar yanlılık olduğu ortaya koyulmuştur. Bu yanlılığın yarışma düzeyine (ulusal, uluslararası, olimpiyatlar) göre farklılaştığını ve yanlılık nedenlerinin de çeşitli nedenlerle açıklandığı (politik nedenler, coğrafi bağlantı, zevk vb.) görülmektedir. ÇDKRM'nin kullanıldığı araştırmalarda puanlamanın modele uyum

sağladığı ve sporcuların yetenek ölçümlerinin kayma becerilerinin anlamlı bir belirleyicisi olduğu ileri sürülmüştür. Genellenebilirlik kuramının kullanıldığı araştırmalarda, genellenebilirlik katsayıları oldukça yüksek bulunmuştur.

BÖLÜM II

YÖNTEM

Bu bölümde araştırmanın türü, araştırma grubu, araştırma verilerinin toplanma şekli ve verilerin analizi üzerinde durulmuştur.

2.1 ARAŞTIRMANIN TÜRÜ

Araştırma, genellenebilirlik kuramının ve ÇDKRM'nin gerçek verilere uygulanmasına, birbirlerine göre avantaj ve sınırlılıklarının ortaya konulmasına, yaklaşımlardan hangisinin daha çok bilgi sağladığının belirlenmesine ve iki kuramın karşılaştırılmasına dayanmaktadır. Dünya Buz Pateni Şampiyonası sonuçlarının genellenebilirlik kuramı ve ÇDKRM ile özelliklerinin belirlenmesine yönelik olduğundan bu yönüyle betimsel bir araştırma niteliği taşımaktadır.

2.2 ÇALIŞMA GRUBU

Bu araştırmanın çalışma grubunu, 2006–2011 yılları arasında yapılan Dünya Buz Pateni Şampiyonalarında tekler (kadınlar ve erkekler) ve çiftler kategorilerinde serbest programda yarışan toplam 397 sporcu ve bu yarışmalarda görev alan toplam 189 hakem oluşturmaktadır.

2.3 ARAŞTIRMA VERİLERİ

Araştırmada, 2006–2011 yılları arasındaki Dünya Buz Pateni Şampiyonalarında serbest programda elde edilen tekler ve çiftler kategorilerindeki hakem değerlendirmeleri kullanılmıştır (<http://www.isureults.com/results>, erişim tarihi: 05.11.2011).

Buz pateninde, teknik puanlar her teknik eleman için ayrı ayrı puanlanır. Her öge (hareket) öncelikle teknik hakem tarafından incelenmektedir. Teknik hakem, teknik

öğenin temel değerini belirlemektedir. Diğer 12 hakem (son düzenlemelerle birlikte 9 hakem) hareketin temel değerine eklenmek üzere GOE (grade of execution) olarak bilinen uygulama puanını vermektedir. Bu, teknik hakem tarafından belirlenen hareketin ne derece iyi gösterildiğini belirlemektedir. Hakemler hareketleri değerlendirmede -3 ve +3 arasında puanlar vermektedir. Puanlama sonucunda, bu hakemler arasından rastgele seçilen 3 tanesi çıkarılmaktadır. Bu işlemin ardından ise geriye kalan puanlar arasından en yüksek ve en düşük puanlar çıkarılmaktadır. Geri kalan hakemlerin puanı GOE'yi belirlemektedir. Hareketin temel değeri ve verilen GOE hareketten alınan puanı oluşturur. Düşme, kaldırışların belirlenen kurallara uymaması, çok uzun sürede yapılması vb. durumlarda ceza puanları (deductions) uygulanmaktadır. Bu çalışmada, 2006–2008 yılları için 12 hakemin, 2009–2011 yılları için 9 hakemin verdiği puanlar analize alınmıştır.

Programlarda hangi teknik öğelerin olacağı ve kaç teknik öğeden oluşacağı yarışmaya ve yarışmanın yapılacağı seviyeye (büyükler, gençler) göre değişir. Büyükler düzeyinde kısa programda, uluslararası yarışmalarda çiftler ve teklere 8 değişik teknik öğe bulunmaktadır. Bu sekiz teknik öğe teklere düzeyinde yarışan patenciler için ISU'nun 310 numaralı kuralında belirtilir. Bütün patenciler kısa programda bir kombinasyon atlayışı, iki solo atlayış, üç dönüş ve iki adım dizisi göstermelidir. Çiftlerde kısa programda ise iki kaldırış, bir yan yana atlayış, bir fırlatmalı atlayış, bir yan yana dönüş, bir birlikte dönüş, bir adım dizisi ve bir ölüm spirali olmalıdır (ISU 313. kural).

Büyükler düzeyinde serbest programda çiftlerde 14 teknik öğe, erkeklerde 13 teknik öğe ve kadınlarda 12 teknik öğe (teknik öğe sayısı yıllara göre farklılık göstermektedir) vardır. Bu teknik öğelerin detayları ISU 520 ve 521. (2008 version) kurallarda verilir. Bunlar; Çiftler kategorisinde 4 kaldırış, 4 atlayış (fırlatmalı ve yan yana), 3 dönüş, 1 ölüm spirali, 1 adım dizisi ve 1 spiral dizisidir. Erkekler 8 atlayış, 3 dönüş ve 2 adım dizisi yapar. Kadınlar ise 7 atlayış, 3 dönüş, 1 adım dizisi ve 1 spiral dizisi gösterir.

Hakemler ayrıca program bileşenlerini de puanlamakta ve 10 üzerinden değerlendirme yapılmaktadır. Bunlar; (1) kayma becerileri (skating skills-SS), (2) geçişler (transitions-TR), (3) performans (performance-PE), (4) koreografi (choreography-CH) ve (5)

yorumlamadır (interpretation IN). Değerlendirmede, her kategori için hakemlerin göz önünde bulundurmaları gereken kurallar vardır (ISU Rules 322.2). 10'luk sistemde değerlendirme yapılmakta ve puanlar 0.25'lik aralıklarla verilmektedir. Bu puanlar programa (kısa-serbest) ve düzeyine göre değişik bir sayı ile çarpılarak program puanını oluşturmaktadır. Büyük kadınlar ve çiftler için bu sayı kısa programda 0.8, serbest programda ise 1.6'dır. Büyük erkeklerde ise kısa program için 1, uzun program için ise 2'dir (ISU, 2010).

Araştırma grubuna ve araştırmada kullanılan verilere ilişkin ayrıntılı bilgiler Tablo 1'de verilmiştir.

Tablo 1. Araştırma grubuna ve araştırma verilerine ilişkin bilgiler

KATEGORİ		N	2006	2007	2008	2009	2010	2011
ERKEK	Serbest Program	Sporcu	24	24	24	24	24	24
		Görev	5	5	5	5	5	5
		Hakem	12	12	12	9	9	9
KADIN	Serbest Program	Sporcu	24	24	23	24	24	24
		Görev	5	5	5	5	5	5
		Hakem	12	12	12	9	9	9
ÇİFT	Serbest Program	Sporcu	20	18	20	20	16	16
		Görev	5	5	5	5	5	5
		Hakem	12	12	12	9	9	9
Toplam Sporcu Sayısı (N)= 397, Toplam Hakem Sayısı=189								

Not: Görev= 5 Program bileşenini (kayma becerileri, geçişler, performans, koreografi, yorumlama) kapsamaktadır.

2.4 VERİLERİ ÇÖZÜMLEME TEKNİKLERİ

Veriler, araştırmanın genellenebilirlik kuramı ve ÇDKRM için kullanılan desenlerine uygun olarak analiz edilmiştir. Genellenebilirlik kuramında, iki değişkenlik kaynaklı (sporcu x görev x hakem) tamamen çaprazlanmış desenler kullanılarak genellenebilirlik ve güvenilirlik katsayıları hesaplanmıştır. Buz pateni puanlama sisteminde, sporculara ilişkin puanlar 9 (veya 12) hakemden alınmaktadır. 2006–2008 yılları arasında yer alan yarışmalarda sporcular 12 hakem tarafından değerlendirilmiştir. Ancak sonraki yıllarda (2009–2011) ekonomik nedenlerden dolayı hakem sayısı 9'a indirilmiştir. 2006–2008 yıllarına ilişkin karar çalışmasında, hakem sayısının 3 azaltıldığı (9'a indirildiği) senaryolara yer verilmiştir. 2009- 2011 yıllarına ilişkin karar çalışmasında ise, hakem sayısının 3 artırıldığı (12'ye çıkarıldığı) senaryolara yer verilmiştir. Bu amaç için EduG (Swiss Society for Research in Education Working Group, 2010) bilgisayar programı kullanılmıştır.

ÇDKRM analizleri FACETS (Linacre, 2007) programı ile yapılmış ve sporcular, hakemler ve görevler için üç değişkenlik kaynaklı desen kullanılmıştır. ÇDKRM'nde model-veri uyumu için standart artıklar ve iç uyum ile dış uyum kareler ortalamaları incelenmiştir. Veri, rasch modeline uyum gösterdiği zaman standartlaştırılmış artıkların %5'inden fazlası ± 2 aralığı, %1'inden fazlası da ± 3 aralığı dışında olmamalıdır (Linacre, 2007). Uyum istatistikleri kareler ortalaması şeklinde rapor edilmektedir. İki tane uyum istatistiği bulunmaktadır: iç uyum ve dış uyum. Dış uyum, uç değerlere duyarlı bir uyum istatistiğidir. Beklenmeyen büyük artıklardan (residuals) etkilenmektedir. Bunun nedeni de gözlenen ve beklenen değerler arasındaki farktır. İç uyum istatistiği ise beklenmeyen küçük artık örüntülerine duyarlıdır (Linacre, 1995). Her iki uyum istatistiğinin 1 olması beklenmektedir. Bu uyum istatistikleri 0 ile sonsuz arasında değerler almaktadır. 1'den küçük olan değerler bağımlılık veya verilerin kısıtlanması, sıkıştırılmasını gösterirken, 1'den büyük değerler ise beklenmeyen değişkenliği, tutarsızlığı göstermektedir (Smith ve Kulikowich, 2004).

İç uyum ve dış uyum kareler ortalaması ile ilgili kabul gören evrensel bir aralık bulunmamaktadır. Verilecek kararın önemi arttıkça iç/dış uyum kareler ortalamasının

1'e yakın olması beklenmektedir (Myford ve Wolfe, 2003). 1'den büyük değerler modelin beklenen değerlerinde daha fazla sapmaları göstermektedir. Diğer hakemlerle aynı fikirde olma veya uyum göstermeye ilişkin psikolojik baskı ve kazanacak buz patenciyi seçmenin önemi düşünüldüğünde iç uyum ve dış uyum kareler ortalamasının 1.2'yi aşması veya .40'dan küçük olması kabul edilmemektedir (Wright ve Linacre, 1994). Program bileşenlerinde (görevler) uyum istatistikleri kabul edilebilir aralıklar içerisindeyse, kolay olan program bileşenleri tüm sporcular ve hakemler için kolay, zor olan program bileşenleri de tüm sporcular ve hakemler için zor olarak yorumlanabilir. Söz konusu sporcular ise; sporculara ilişkin uyum istatistikleri istenilen sınırlar içerisinde değilse, o sporcunun yetenek kestirimine güvenemeyiz. Hakemler için uyum istatistikleri ise hakemin tüm sporcuları ve tüm program bileşenlerini puanlamadaki tutarlılığına ilişkin bilgi sağlamaktadır. Uyum istatistikleri kabul edilebilir aralıklarda değilse, hakemlerin yetenekli sporcuları kolay görevlerde düşük puanladıkları veya yetenek düzeyi daha düşük sporcuları zor görevlerde yüksek puanladıkları şeklinde yorumlanabilir (Myford ve Wolfe, 2003).

Sporcuların yetenekleri ve uygunluk istatistikleri ile hakemlerin katılık/cömertlik istatistiklerinin belirlenmesi amacı ile tamamı aynı X^2 , rastgele normal X^2 değerleri, güvenilirlik ve ayırma oranı indeksleri incelenmiştir. ÇDKRM'nda iki farklı güvenilirlik katsayısı bulunmaktadır. Bunlar; ayırma indeksinin güvenilirliği ve ayırma oranıdır. Bu iki istatistik farklı ölçeklerde rapor edilmektedir. Aynı bilgiler kullanılarak hesaplanmakta ve benzer sonuçlara ulaşmaktadır. ÇDKRM ile kestirilen ölçümlerin değişkenliğine ilişkin bilgi vermektedirler. Ayırma indeksinin güvenilirliği 0 ile 1.0 arasında değer alırken, ayırma oranı ise 0 ile sonsuz arasında değer almaktadır. Her bir değişkenlik kaynağı için bu değerler hesaplanmaktadır. Ancak bu değerlerin yorumları her bir değişkenlik kaynağı için farklılaşmaktadır (Bond ve Fox, 2007).

Birey değişkenlik kaynağı söz konusu olduğunda; ayırma indeksinin güvenilirliği ve ayırma oranının yüksek olması arzu edilen bir durumdur. Birey değişkenlik kaynağı için ayırma indeksinin güvenilirliği Cronbach alfa ile benzemektedir (Myford ve Wolfe, 2003; Sudweeks, Reeve ve Bradshaw, 2005). Gerçek varyansın gözlenen varyansa oranını temsil etmektedir. Bire yakın değerler yüksek güvenilirliği göstermektedir. Birey

değişkenlik kaynağı söz konusu olduğunda bu değerlerin düşük olması, merkezi eğilim hatasını göstermektedir. Puanlayıcıların bireylerin performansını, söz konusu görevde ayırt edemediğini göstermektedir (Myford ve Wolfe, 2003; Sudweeks, Reeve ve Bradshaw, 2005).

G- Kuramı ve ÇDKRM'ne göre sporcu, hakem ve sporcuhakem değişkenlik boyutlarında elde edilen parametrelerin benzerlik ve farklılıkları belirlenmiştir. Bunların belirlenmesinde, G-kuramı için toplam varyansı açıklama yüzdeleri, G ve Phi katsayıları incelenirken, ÇDKRM için ayırma indeksi güvenilirliği, ayırma oranı, tamamı aynı kare ve sporcuhakem yanlılık yüzdeleri incelenmiştir.

BÖLÜM III

BULGULAR VE YORUM

Bu bölümde, araştırmanın problem cümlesi ve alt problemlerin sırasına göre elde edilen bulgular ve bulgulara ilişkin tartışmalar yer almaktadır.

Alt Problem I: 2006–2011 yıllarındaki Dünya Buz Pateni Şampiyonası sonuçlarına ilişkin G-çalışması sonuçları (varyans bileşenleri ve toplam varyansı açıklama yüzdeleri):

Buz pateni performansına ilişkin genellenebilirlik çalışması için 2006–2011 yılları arasında yapılan Dünya Buz Pateni Şampiyonlarından alınan tekler (kadınlar ve erkekler) ve çiftler yarışmalarına ait hakem değerlendirmelerine (müsabaka sonuçlarına) tamamen çaprazlanmış sxg_xh (s=sporcu, g=görev, h=hakem) deseni uygulanmıştır.

2006–2011 yıllarında farklı kategorilerde yapılan G-çalışması için kestirilen varyans bileşenleri ve varyansı açıklama yüzdeleri s, g, h ana etkileri ile sg, sh, gh ve sgh ortak etkileri erkekler serbest program için Tablo 2’de, kadınlar serbest program için Tablo 3’te ve çiftler serbest program için Tablo 4’te verilmiştir (Tüm bu analizlere ilişkin kareler ortalaması, serbestlik derecesi, kareler toplamı, varyans ve varyansı açıklama yüzdeleri EK 1’de ayrıntılı olarak bulunmaktadır).

Tablo 2. 2006-2011 yıllarındaki Dünya Buz Pateni Şampiyonası erkekler serbest program sonuçlarına ilişkin varyans bileşenleri ve toplam varyansı açıklama yüzdeleri

Varyansın Kaynağı	2006		2007		2008		2009		2010		2011	
	Varyans	%	Varyans	%	Varyans	%	Varyans	%	Varyans	%	Varyans	%
S	39.128	77.9	69.139	82.6	95.254	81.7	79.156	78.8	106.226	79.1	86.727	76.6
G	0.491	1.0	2.581	3.1	2.986	2.6	2.761	2.7	3.301	2.5	3.644	3.2
H	0.196	0.4	0.852	1.0	2.063	1.8	1.075	1.1	-0.494	0.0	0.576	0.5
SG	0.708	1.4	0.422	0.5	0.955	0.8	0.904	0.9	0.948	0.7	1.284	1.1
SH	5.187	10.3	8.200	9.8	10.296	8.8	12.481	12.4	14.953	11.1	12.640	11.2
GH	0.634	1.3	0.258	0.3	1.045	0.9	0.748	0.7	0.037	0.0	-0.075	0.0
SGH	3.854	7.7	2.297	2.7	3.986	3.4	3.320	3.3	8.796	6.6	8.399	7.4

S=Sporcu, G= Görev, H= Hakem

Tablo 2’de verilen 2006–2011 Dünya Buz Pateni Şampiyonası erkekler serbest program sonuçlarına ilişkin varyans bileşenleri ve toplam varyansı açıklama yüzdeleri incelendiğinde, 2006–2008 yıllarında erkekler serbest programda sporcu (s) ana etkisi için kestirilen varyans bileşenlerinin toplam varyansın yıllara göre sırasıyla %77.9, %82.6 ve %81.7’sini açıkladığı görülmektedir. Sporcular için kestirilen varyans bileşeninin, toplam varyans içerisindeki en yüksek bileşen olduğu görülmektedir. Sporcular için kestirilen varyansın, toplam varyans içerisindeki oranının en yüksek olması genellenebilirlik çalışmalarında arzu edilen bir durumdur. G-çalışmalarında ölçme nesnesinin (sporcu) ana etkisi evren puanı varyansı olarak değerlendirilmektedir (Brennan, 2001). Bu da, performans açısından bireyler arası farklılıkları göstermektedir. Yani, ölçülen özellik açısından bireyler arası farklılıkların belirlenebildiğini söyleyebiliriz.

Genellenebilirlik çalışması sonuçlarına göre, sporcuhakem etkileşiminin etkisi için kestirilen varyans bileşenleri yıllara göre sırasıyla toplam varyansın %10.3, %9.8 ve %8.8’ini açıklamaktadır. Bu da hakemlerin farklı sporcuları farklı katılık/cömertlikte puanladığının bir göstergesi olarak yorumlanabilir (Huang ve Foote, 2011). 2006–2008 yıllarında erkekler serbest programda hakem (h) ana etkisi için kestirilen varyans bileşenlerinin toplam varyansın yıllara göre sırasıyla %0.4, %1.0 ve %1.8’ini açıkladığı görülmektedir. Hakemler için kestirilen varyans bileşeninin, toplam varyans içerisindeki yerinin düşük olduğu görülmektedir. Bu bulgu, hakemlerin katılık/cömertlik düzeyinin tutarlı olduğunu göstermektedir. Huang ve Foote (2011), 2004 Dünya Buz Pateni Şampiyonası verileri ile yaptıkları çalışmalarında tek erkeklerde, sporcu ana etkisinin %66.2 ile ilk sırada yer aldığını bulmuştur. Bu da elde edilen bulguları desteklemektedir. Hakem ana etkisinden elde edilen varyansın, toplam varyansın %7.90’ını açıkladığını bulmuşlardır. Bu bulgu, araştırma bulguları ile benzer nitelikte değildir. Bunun nedeni ise, 2006 yılında Buz Pateni performansının değerlendirilmesinde yapılan değişiklikler olabilir.

Erkekler serbest programda, 2006–2008 yıllarında yapılan yarışmalarda elde edilen sonuçlarla ilgili genellenebilirlik çalışmasında kestirilen varyans bileşenleri incelendiğinde, toplam varyans içerisinde en yüksek üçüncü etkiye

sporcu görev hakem etkileşiminin sahip olduğu görülmektedir. Bu, etkileşime ek olarak ölçme hatalarının açıklanamayan kısmını da temsil etmektedir ve düşük olması beklenmektedir. Buna göre, hakemlerin sporcuları bir görevden diğerine farklı katılık/cömertlik düzeyinde sıraladıklarını, yani az da olsa tutarsızlık olduğunu göstermektedir.

2009–2011 yıllarında erkekler serbest programda sporcu (s) ana etkisi için kestirilen varyans bileşenlerinin toplam varyansın yıllara göre sırasıyla %78.8, %79.1 ve %76.6'sını açıkladığı görülmektedir. Sporcular için kestirilen varyans bileşeninin, toplam varyans içerisindeki en yüksek bileşen olduğu görülmektedir. Sporcular için kestirilen varyansın, toplam varyans içerisindeki oranının en yüksek olması genellenebilirlik çalışmalarında arzu edilen bir durumdur. Bu da, performans açısından bireyler arası farklılıkları göstermektedir.

Genellenebilirlik çalışması sonuçlarına göre, sporcu hakem etkileşiminin etkisi için kestirilen varyans bileşenleri yıllara göre sırasıyla toplam varyansın %12.4, %11.1 ve %11.2'sini açıklamaktadır. Bu da hakemlerin sporcuları, farklı katılık/cömertlikte puanladığının bir göstergesi olarak yorumlanabilir. Sporcu görev hakem etkileşimi için kestirilen varyans bileşenleri yıllara göre sırasıyla toplam varyansın %3.3, %6.6 ve %7.4'ünü açıkladığı görülmektedir. Bu da hakemlerin görevleri puanlarken farklı puanlama yaptığını göstermektedir. Ayrıca, sistematik ve sistematik olmayan hata kaynaklarının oranını da göstermektedir. 2009–2011 yıllarında, sporcu hakem etkileşiminin toplam varyansı açıklama yüzdesi 2006–2008 yıllarına göre daha yüksek bulunmuştur. Bu da hakem sayısındaki değişime bağlı olabilir.

Tablo 3'te 2006–2011 Dünya Buz Pateni Şampiyonası kadınlar serbest program sonuçlarına ilişkin varyans bileşenleri ve toplam varyansı açıklama yüzdeleri verilmiştir.

Tablo 3. 2006–2011 yıllarındaki Dünya Buz Pateni Şampiyonası kadınlar serbest program sonuçlarına ilişkin varyans bileşenleri ve toplam varyansı açıklama yüzdeleri

Varyansın Kaynağı	2006		2007		2008		2009		2010		2011	
	Varyans	%	Varyans	%	Varyans	%	Varyans	%	Varyans	%	Varyans	%
S	91.361	86.1	102.767	83.8	89.036	80.3	166.538	85.0	110.690	77.5	116.577	88.7
G	1.787	1.7	3.035	2.5	2.909	2.6	2.687	1.4	3.647	2.6	1.792	1.4
H	1.285	1.2	0.703	0.6	4.103	3.7	6.600	3.4	0.087	0.1	-0.081	0.0
SG	0.330	0.3	0.486	0.4	0.539	0.5	1.036	0.5	1.150	0.8	0.329	0.3
SH	8.538	8.0	10.885	8.9	9.812	8.8	13.250	6.8	20.529	14.4	9.670	7.4
GH	0.575	0.5	0.600	0.5	1.215	1.1	1.587	0.8	-0.037	0.0	0.029	0.0
SGH	2.290	2.2	4.211	3.4	3.329	3.0	4.258	2.2	6.774	4.7	3.064	2.3

S= Sporcu, G= Görev, H=Hakem

Tablo 3 incelendiğinde, 2006–2008 yıllarında kadınlar serbest programda sporcu (s) ana etkisi için kestirilen varyans bileşenlerinin toplam varyansı açıklamadaki yerinin en büyük olduğu görülmektedir. Kadınlar serbest programda sporcu ana etkisinin toplam varyansı açıklama oranı yıllara göre sırasıyla %86.1, %83.8 ve %80.3 olarak bulunmuştur. Bu, arzu edilen bir durumdur. Ölçülen özellik açısından bireyler arası farklılıkların belirlenebildiğini söyleyebiliriz.

İkinci en yüksek varyansa sahip olan bileşen sporcuhakem etkileşimidir. Kadınlar serbest programda, sporcuhakem etkileşiminin toplam varyansı açıklama oranı yıllara göre sırasıyla %8.0, %8.9 ve %8.8 olarak bulunmuştur. Bu da hakemlerin farklı sporcuları farklı katılık/cömertlikte puanladığının bir göstergesi olarak yorumlanabilir (Huang ve Foote, 2011). Huang ve Foote (2011), 2004 Dünya Buz Pateni Şampiyonası verileri ile yaptıkları çalışmalarında kadınlarda sporcu ana etkisinin %66.2 ile ilk sırada yer aldığını, hakem ana etkisinden elde edilen varyansın ise toplam varyansın %9.80'nini açıkladığını bulmuşlardır.

Genellenebilirlik sonuçlarına bakıldığında, sporcugörevhake bileşeninin varyansının, toplam varyansı açıklamadaki payı üçüncü sırada yer almaktadır. Kadınlar serbest programda toplam varyansı açıklama yüzdesi yıllara göre %2.2, %3.4 ve %3.0'dır. Bu da hakemlerin sporcuları farklı görevlerde, çok az tutarsızlıkla puanladıkları anlamına gelmektedir. Bunun yanında sistematik ve sistematik olmayan hata kaynaklarını da içermektedir. Üç değişkenlik kaynağının etkileşiminden elde edilen bu varyansın sıfır (0)'a yakın olması istenilen bir durum olmakla birlikte, ölçme hatasının az olduğunu göstermektedir.

2009–2011 Dünya Buz Pateni Şampiyonası kadınlar serbest program sonuçlarına ilişkin genellenebilirlik analizleri incelendiğinde; sporcu ana etkisinin toplam varyansı açıklamada en büyük paya sahip olduğu görülmektedir. Kadınlar serbest programda toplam varyansı açıklama yüzdeleri yıllara göre sırasıyla; %85.0, %77.5 ve %88.7'dir. Sporcular için kestirilen varyansın, toplam varyans içerisindeki oranının en yüksek

olması genellenebilirlik çalışmalarında arzu edilen bir durumdur. Bu da, performans açısından bireyler arası farklılıkların belirlenebildiği anlamına gelmektedir.

İkinci en yüksek varyansa sahip bileşen sporcuhakem etkileşimi olarak bulunmuştur. Sporcuhakem etkileşiminin toplam varyansı açıklama yüzdeleri yıllara göre sırasıyla; %6.8, %14.4 ve %7.4 olarak bulunmuştur. Hakemler farklı sporcuları farklı katılık/cömertlik düzeyinde puanlamaktadırlar. Üçüncü sırada ise sporcugörevhakem etkileşimi yer almaktadır. Bu da her üç bileşenin ortak etkisini ve hatayı içermektedir.

Tablo 4'te 2006–2011 Dünya Buz Pateni Şampiyonası çiftler serbest program sonuçlarına ilişkin varyans bileşenleri ve toplam varyansı açıklama yüzdeleri verilmiştir.

Tablo 4. 2006–2011 yıllarındaki Dünya Buz Pateni Şampiyonası çiftler serbest program sonuçlarına ilişkin varyans bileşenleri ve toplam varyansı açıklama yüzdeleri

Varyansın Kaynağı	2006		2007		2008		2009		2010		2011	
	Varyans	%	Varyans	%	Varyans	%	Varyans	%	Varyans	%	Varyans	%
S	202.373	94.4	133.608	91.6	240.670	92.2	172.870	90.8	176.514	91.2	194.674	89.4
G	1.540	0.7	1.382	0.9	1.656	0.6	1.115	0.6	1.872	1.0	2.103	1.0
H	0.755	0.4	-0.066	0.0	3.853	1.5	1.493	0.8	0.241	0.1	-0.486	0.0
SG	0.307	0.1	0.267	0.2	0.631	0.2	0.526	0.3	0.319	0.2	0.648	0.3
SH	6.067	2.8	8.029	5.5	11.022	4.2	10.152	5.3	11.152	5.8	16.161	7.4
GH	0.520	0.2	0.279	0.2	0.424	0.2	0.796	0.4	-0.074	0.0	-0.019	0.0
SGH	2.763	1.3	2.342	1.6	2.859	1.1	3.458	1.8	3.459	1.8	4.255	2.0

S= Sporcu, G= Görev, H= Hakem

Tablo 4'te verilen 2006–2011 Dünya Buz Pateni Şampiyonası çiftler serbest program sonuçlarına ilişkin varyans bileşenleri ve toplam varyansı açıklama yüzdeleri incelendiğinde, 2006–2011 yıllarında çiftler serbest programda sporcu (s) ana etkisi için kestirilen varyans bileşenlerinin toplam varyansın yıllara göre sırasıyla %94.4, %91.6, %92.2, %90.8, %91.2 ve %89.4'ünü açıkladığı görülmektedir. Sporcular için kestirilen varyans bileşeninin, toplam varyans içerisindeki en yüksek bileşen olduğu görülmektedir. Bu da, performans açısından bireyler arası farklılıkları göstermektedir. Genellenebilirlik çalışması sonuçlarına göre, sporcuhakem etkileşiminin etkisi için kestirilen varyans bileşenleri yıllara göre sırasıyla toplam varyansın %2.8, %5.5, %4.2, %5.3, %5.8 ve %7.4'ünü açıklamaktadır. Bu da hakemlerin farklı sporcuları farklı katılık/cömertlikte puanladığının bir göstergesi olarak yorumlanabilir (Huang ve Foote, 2011). Çiftlerde elde edilen sporcuhakem etkileşimleri, tekler programında elde edilenlere göre daha düşük bulunmuştur. Bunun nedeni, 2002 Kış Olimpiyatlarında, çiftler yarışmasında ortaya çıkan skandal olabilir. Huang ve Foote (2011), 2004 Dünya Buz Pateni Şampiyonası verileri ile yaptıkları çalışmalarında çiftlerde sporcu ana etkisinin %89.63 gibi yüksek bir varyans açıklama yüzdesine sahip olduğunu, hakem ana etkisinden elde edilen varyansın, toplam varyansın %3.59'unu açıkladığını bulmuşlardır. Hellsten, Henderson ve Klinger (2000) buz pateninde sporcu ana etkisini %67 ile %91 arasında bulmuşlardır. Bu bulgular araştırmada elde edilen bulguları desteklemektedir.

Alt Problem II: 2006–2008 ve 2009–2011 yılları arasında elde edilen verilerde orijinal hakem sayısı ile hakem sayısının artırılıp azaltılması ile yapılan genellenebilirlik ve karar çalışması sonuçları (G ve Phi katsayıları):

Dünya Buz Pateni Şampiyonalarından elde edilen sonuçlar üzerinden yapılan karar çalışmaları sonucunda mutlak ve bağıl hata varyansları ile G ve Phi katsayıları kestirilmiştir. 2006–2008 yılları arasında yer alan yarışmalarda sporcular 12 hakem tarafından değerlendirilmiştir. Ancak sonraki yıllarda (2009–2011) ekonomik nedenlerden dolayı hakem sayısı 9'a indirilmiştir. 2006–2008 yıllarında orijinal hakem sayısının 12 olduğu genellenebilirlik çalışması ve hakem sayısının 3 azaltıldığı (9'a

indirildiği) karar çalışmalarına ilişkin senaryolara yer verilmiştir. Bu karar çalışmalarına ilişkin sonuçlar Tablo 5'te yer almaktadır.

Tablo 5. 2006–2008 yıllarında 12 hakem ile yapılan genellenebilirlik ve 9 hakem ile yapılan karar çalışmalarından elde edilen G ve Phi katsayıları ile hata varyansları

		2006		2007		2008	
		12	9	12	9	12	9
Erkekler Serbest Program	G	.98 (.64)	.98 (.80)	.99 (.81)	.99 (1.05)	.99 (1.12)	.99 (1.42)
	Phi (Φ)	.98 (.76)	.98 (.94)	.98 (1.39)	.98 (1.66)	.98 (1.90)	.98 (2.27)
Kadınlar Serbest Program	G	.99 (.82)	.99 (1.07)	.99 (1.07)	.99 (1.40)	.99 (.98)	.99 (1.27)
	Phi (Φ)	.99 (1.29)	.98 (1.58)	.98 (1.75)	.98 (2.09)	.98 (1.93)	.97 (1.13)
Çiftler Serbest Program	G	1.00 (.61)	1.00 (.79)	.99 (.76)	.99 (.99)	1.00 (1.09)	.99 (1.41)
	Phi (Φ)	1.00 (.99)	.99 (1.20)	.99 (1.04)	.99 (1.28)	.99 (1.75)	.99 (2.18)

NOT: () içerisinde verilen değerler hata varyanslarıdır.

Tablo 5 incelendiğinde, hakem sayısının 12 olduğu yarışmalarda G-katsayısının .98 ve 1.00 arasında, phi katsayısının ise .98 ve 1.00 arasında değiştiği görülmektedir. Karar çalışmasında hakem sayısının 9'a indirilmesi sonucunda G (.98 ve 1.00) ve phi (.98 ve .99) katsayılarında düşüş meydana gelmemiş, bağıl ve mutlak hata varyanslarında da çok az bir artış meydana gelmiştir. Hakem sayısının düşürülmesi genellenebilirlik ve güvenilirlik katsayılarında değişikliğe yol açmamaktadır.

2009–2011 yıllarında orijinal hakem sayısının 9 olduğu genellenebilirlik çalışması ve hakem sayısının 3 artırıldığı (12'ye çıkarıldığı) karar çalışması senaryolarına ilişkin sonuçlar Tablo 6'da yer almaktadır.

Tablo 6. 2009–2011 yıllarında 9 hakem ile yapılan genellenebilirlik ve 12 hakem ile yapılan karar çalışmalarından elde edilen G ve Phi katsayıları ile hata varyansları

		2009		2010		2011	
		9	12	9	12	9	12
Erkekler Serbest Program	G	.98 (1.64)	.98 (1.28)	.98 (2.05)	.99 (1.58)	.98 (1.85)	.98 (1.45)
	Phi (Φ)	.97 (2.33)	.98 (1.93)	.98 (2.71)	.98 (2.24)	.97 (2.64)	.97 (2.23)
Kadınlar Serbest Program	G	.99 (1.77)	.99 (1.38)	.98 (2.66)	.98 (2.05)	.99 (1.21)	.99 (.92)
	Phi (Φ)	.98 (3.08)	.99 (2.49)	.97 (3.40)	.98 (2.79)	.99 (1.57)	.99 (1.28)
Çiftler Serbest Program	G	.99 (1.31)	.99 (1.01)	.99 (1.38)	.99 (1.05)	.99 (2.02)	.99 (1.55)
	Phi (Φ)	.99 (1.72)	.99 (1.37)	.99 (1.78)	.99 (1.45)	.99 (2.44)	.99 (1.97)

NOT: () içerisinde verilen değerler hata varyanslarıdır.

Tablo 6 incelendiğinde, 2009–2011 yıllarında elde edilen G ve phi katsayılarının yüksek olduğu görülmektedir. Hakem sayısının 9 olduğu yarışmalarda G-katsayısının .98 ve .99 arasında, phi katsayısının ise yine .97 ve .99 arasında değiştiği görülmektedir. Karar çalışmaları açısından incelendiğinde ise 2009–2011 verilerinde hakem sayısının 12'ye çıkarılması artışa neden olmamaktadır. Bunun yanında, 2009–2011 yıllarında Dünya Buz Pateni Şampiyonalarında hakem sayısında azaltmaya gidilmesinin, genellenebilirlik ve güvenilirlik (Φ) katsayılarını olumsuz etkilemediğini söyleyebiliriz.

Huang ve Foote (2011), 2004 Dünya Buz Pateni Şampiyonası verileri ile yaptıkları çalışmada G-katsayılarını kadınlarda ve erkeklerde .93, çiftlerde ise .99 olarak bulmuşlardır. Saeki ve Fan (1999) kış olimpiyatlarında buz pateninden elde edilen puanlarla yaptığı çalışmasında genellenebilirlik katsayılarının oldukça yüksek (.98 ve .99) olduğunu bulmuştur. Hellsten, Henderson ve Klinger (2000) buz pateninde genellenbilirlik katsayılarını .95 ile .99 arasında bulmuşlardır. Bu bulgular da buz pateninde yapılan değerlendirmelerin genellenebilirlik katsayılarının yüksek olduğunu ve yapılan değerlendirmenin güvenilir olduğunu göstermektedir.

Buz pateni puanlama sisteminde, sporculara ilişkin puanlar 9 (veya 12) hakemden alınmaktadır. Burada seçilen 9 (veya 12) hakem, sahip oldukları vasıflardan dolayı tüm dünyadaki nitelikli hakemler popülasyonundan seçilen küçük bir örneklem olarak görülebilir. Bunun yanında, G-kuramı açısından bakıldığında, 9 (veya 12) hakemden elde edilen puanlar, tüm olası hakemlerden elde edilen puanları temsil etmektedir (Furr ve Bacharach, 2008). Böylece, yapılan karar çalışmalarına ve genellenebilirlik çalışmalarına ilişkin sonuçlar incelendiğinde, hakem sayısının karar çalışmaları ile artırılıp azaltılması ile elde edilen bulgular, orijinal hakem sayılarından elde edilen bulgularla benzerlik göstermektedir. Bu da, karar çalışmalarının geçerliğine ilişkin bir bulgu olarak yorumlanabilir.

Alt Problem III: 2006- 2008 yılları arasında buz pateni performansının ölçülmesinde 12 hakemin kullanıldığı yarışmalar (model veri uygunluğu, sporcuların yetenekleri ve uygunluk istatistikleri, hakemlerin katılık/cömertlik istatistikleri, hakemxsporcu yanlılığı) için ÇDKRM analizi sonuçları:

Dünya Buz Pateni Şampiyonalarından (2006–2011) elde edilen hakem değerlendirmelerine ilişkin verilerin çok değişkenlik kaynaklı Rasch modeli analizleri için FACETS bilgisayar programı kullanılmıştır. Sporcu, görev ve hakemlerden oluşan üç değişkenlik kaynaklı desen kullanılmıştır. Bunun sonucunda elde edilen model veri uygunluğu, sporcuların yetenekleri, görevlerin güçlüğü ve hakemlerin katılık cömertlik istatistikleri sırasıyla verilmiştir.

Dünya Buz Pateni Şampiyonalarında, 2006–2008 yıllarında erkekler, kadınlar ve çiftler serbest program yarışmalarında elde edilen sonuçlara ilişkin verilerin modelle uyumunu test etmek amacı ile ÇDKRM analizleri sonucunda elde edilen standartlaştırılmış artık değerleri incelenmiştir. Buna ilişkin sonuçlar Tablo 7’de verilmiştir.

Tablo 7. 2006–2008 Dünya Buz Pateni Şampiyonası serbest program değerlendirmelerine ilişkin ÇDKRM analizleri sonucunda elde edilen standart artık yüzdeleri, sporcu, görev ve hakem değişkenlik kaynaklarına göre uyumsuzluk yüzdeleri

Yıl	Bireyler	Standart Artık (%)	Uyumsuzluk Sporcular	Uyumsuzluk Görevler	Uyumsuzluk Hakemler
2006	Erkek	21 (%1.5)	n=7 (%29.2) 2,3,6,10,17,18,23	n=1 Geçişler	n=2 4,6
	Kadın	5 (%0.3)	n=5 (%20) 5,8,12,16,18	—	n=4 2,6,7,8
	Çift	2 (%0.2)	n=4 (%20) 1,6,13,19	—	n=3 2,4,5
2007	Erkek	7 (%0.5)	n=6 (%25) 2,13,17,19,22,24	—	n=2 5,11
	Kadın	12 (%0.8)	n=9 (%37.5) 2,3,5,12,13,14,16,20,22	n=1 Yorumlama	n=6 1,2,5,8,10,12
	Çift	8 (%0.7)	n=2 (%11.1) 4,18	—	n=2 5,7
2008	Erkek	10 (%0.7)	n=7 (%29.2) 2,8,11,12,18,20,23	—	n=3 2,5,6
	Kadın	6 (%0.4)	n=7 (%30.4) 1,10,12,17,18,19,21	—	n=3 2,3,6
	Çift	4 (%0.3)	n=5 (%25) 2,4,9,10,15	—	n=3 3,6,7

Tablo 7 incelendiğinde ± 3 aralığı dışında olan standartlaştırılmış artık değerlerinin oranının %0.2 ile %1.5 arasında değiştiği bulunmuştur. Sadece 2006 erkekler serbest programda elde edilen veriye ilişkin ± 3 aralığı dışında olan artıkların %1.5 olduğu görülmektedir. Ancak çok aşırı bir sapma olmadığından dolayı model veri uyumunun sağlandığı söylenebilir (Looney, 2012). Ayrıca, Tablo 7’de sporcular, görevler ve hakemler için uyumsuzluk sayıları ve uyumsuzluk gösteren sporcular, görevler ve hakemler belirtilmiştir. Bu sayılar belirlenirken iç uyum ve dış uyum kareler ortalamasının 0.40’dan küçük veya 1.2’den büyük olduğu durumlar dikkate alınmıştır. Görevler açısından incelendiğinde 2006 erkekler serbest program ve 2007 kadınlar serbest programda tek bir görev açısından uyumsuzluk belirlenmiştir. Bu uyumsuzluğun, birinin yorumlamada, diğerinin ise geçişlerde olduğu tespit edilmiştir. Looney (2012) Olimpiyat Buz Pateni verileri ile yaptığı çalışmada, geçişlerin %26.4, yorumlama ve koreografinin de %17 ile en fazla hatalı değerlendirme yapılan kategoriler olduğunu belirlemiştir.

Hakemler açısından incelendiğinde ise; uyumsuzluk gösteren hakem sayısının 2 ile 6 arasında değiştiği görülmektedir. Hakemler için uyum istatistikleri ise hakemin tüm sporcuları ve tüm program bileşenlerini puanlamadaki tutarlılığına ilişkin bilgi sağladığından tabloda belirtilen hakemlerin kendi içlerindeki puanlamalarının tutarlı olmadığı söylenebilir. Ancak uyum istatistikleri, hakemlerin tüm sporcuları puanlamada tutarlı olup olmadığını belirlemede yeterli olmadığından yanlılık analizi de yapılmıştır. Buna göre sporcuxhakem yanlılık oranları hesaplanmıştır. Buna göre; 2006 erkekler serbest programda %10.8, 2006 kadınlar serbest programda %7.7, 2006 çiftler serbest programda %2.8; 2007 erkekler serbest programda %9.6, 2007 kadınlar serbest programda %8.6, 2007 çiftler serbest programda %5.3; 2008 erkekler serbest programda %8.5, 2008 kadınlar serbest programda %8.3, 2008 çiftler serbest programda %3.9 düzeyinde hakemler sporcuları farklı puanlamışlardır.

Sporcular açısından incelendiğinde ise; uyumsuzluk gösteren sporcu sayısının 2 ile 9 arasında değiştiği görülmektedir. Bu da, bu sporcuların performansındaki değişkenliğin beklenen modelden anlamlı olarak farklı olduğunu göstermektedir. Genel olarak uyum istatistikleri 2'nin altında olduğundan dolayı, bu sonuçların genel sonuçları bozmadığı düşünülmektedir (Sudweeks, Reeve ve Bradshaw, 2005). Sporculara ilişkin uyum istatistiklerinin istenilen sınırlarda olmaması o sporcuya ilişkin yetenek kestirimine güvenilmeyeceğini ifade etmektedir.

Tablo 8 ÇDKRM analizleri sonucunda, 2006–2008 yılları arasında erkekler, kadınlar ve çiftler serbest program yarışmalarına ilişkin en düşük ve en yüksek logit değerlerini, logit ortalama ve standart sapmalarını, iç uyum ve dış uyum kareler ortalamalarının ve standart sapma değerlerini vermektedir. Tabloda her bir değişkenlik kaynağı için bu değerler ayrı ayrı verilmiştir (Sporcuların yeteneklerine, görevlerin güçlüğüne ve hakemlerin katılımına ilişkin bireysel düzeydeki logit, iç uyum ve dış uyum değerleri EK-2'de verilmiştir).

Tablo 8. 2006–2008 Dünya Buz Pateni Şampiyonaları serbest program değerlendirmelerine ilişkin ÇDKRM değişkenlik kaynaklarına ait logit minimum/maksimum, logit, iç uyum ve dış uyum ortalama ve standart sapma değerleri.

		Logit min-max	Logit Ort.± Ss	İç uyum Ort ±Ss	Dış uyum Ort± Ss
2006 Erkek	Sporcular	(-2.55)-(3.01)	-.04±1.57	.99±.36	.99±.36
	Görevler	(-.29)-(.27)	.00±.18	.99±.12	.99±.13
	Hakemler	(-.35)-(.35)	.00±.20	.99±.73	.99±.72
2006 Kadın	Sporcular	(-3.36)-(3.35)	-.16±2.13	.98±.32	.98±.31
	Görevler	(-.52)-(.38)	.00±.29	.98±.09	.98±.08
	Hakemler	(-.68)-(.52)	.00±.30	.98±.44	.98±.44
2006 Çift	Sporcular	(-6.35)-(6.78)	.83±4.15	.96±.26	.96±.26
	Görevler	(-.50)-(.45)	.00±.32	.97±.12	.96±.12
	Hakemler	(-.51)-(.63)	.00±.30	.97±.33	.96±.31
2007 Erkek	Sporcular	(-3.07)-(4.18)	.36±2.14	.99±.27	.99±.27
	Görevler	(-.63)-(.50)	.00±.36	1.00±.13	.99±.13
	Hakemler	(-.41)-(.47)	.00±.27	1.00±.44	.99±.44
2007 Kadın	Sporcular	(-2.53)-(3.47)	.22±1.68	.98±.41	.98±.40
	Görevler	(-.45)-(.40)	.00±.27	.98±.14	.98±.13
	Hakemler	(-.39)-(.30)	.00±.19	.97±.54	.98±.55
2007 Çift	Sporcular	(-4.49)-(5.93)	-.69±2.69	.90±.32	.90±.32
	Görevler	(-.37)-(.41)	.00±.26	.91±.13	.90±.13
	Hakemler	(-.24)-(.28)	.00±.15	.91±.55	.90±.55
2008 Erkek	Sporcular	(-2.35)-(3.79)	.52±1.73	1.00±.36	.99±.35
	Görevler	(-.48)-(.36)	.00±.27	.98±.11	.99±.12
	Hakemler	(-.62)-(.40)	.00±.27	.98±.50	.99±.51
2008 Kadın	Sporcular	(-2.71)-(3.47)	.11±1.85	.99±.37	.98±.37
	Görevler	(-.54)-(.34)	.00±.30	.98±.08	.98±.07
	Hakemler	(-.66)-(.61)	.00±.40	.98±.43	.98±.44
2008 Çift	Sporcular	(-5.02)-(5.09)	.49±2.74	.97±.31	.97±.30
	Görevler	(-.36)-(.31)	.00±.22	.97±.11	.97±.10
	Hakemler	(-.88)-(.79)	.00±.38	.97±.41	.97±.40

Not: Ortalamanın +1 ve -1 standart sapma aralığı verilmiştir.

Tablo 8 incelendiğinde, iç uyum kareler ortalaması ve dış uyum kareler ortalamasına ilişkin ortalamaların tüm gruplarda ve tüm değişkenlik kaynaklarında 1'e yakın olduğu görülmektedir. Bu da verilerin model ile uyumlarının iyi olduğuna işaret etmektedir. Yani başka bir deyişle, veriler ölçme için uygundur (Linacre, 2007).

Tablo 9'da ise sporcular için ayırma indeksi oranı, ayırma indeksi güvenilirliği, sabit etki ve rastgele etki hipotezine ilişkin X^2 testi sonuçları yer almaktadır.

Tablo 9. 2006–2008 Dünya Buz Pateni Şampiyonası serbest programa ilişkin sporcular için ayırma oranı, güvenilirlik ve X^2 testleri.

	RMSEA	Ss	G	Güvenirlik	Tamamı aynı X^2	Sd	p	Rastgele Normal X^2	Sd	p
2006 Erkek	.11	1.56	14.60	1.00	4798.9	23	.00	22.9	22	.41
2006 Kadın	.10	2.13	21.24	1.00	10677.6	23	.00	23.0	22	.40
2006 Çift	.11	4.15	37.75	1.00	29291.6	19	.00	19.0	18	.39
2007 Erkek	.10	2.14	20.66	1.00	11166.8	23	.00	23.0	22	.40
2007 Kadın	.09	1.68	19.61	1.00	9540.4	23	.00	22.9	22	.40
2007 Çift	.10	2.69	26.49	1.00	11289.1	17	.00	17.0	16	.39
2008 Erkek	.09	1.73	19.69	1.00	9251.6	23	.00	22.9	22	.41
2008 Kadın	.09	1.84	20.40	1.00	10608.5	22	.00	22.0	21	.40
2008 Çift	.09	2.74	30.29	1.00	17588.2	19	.00	19.0	18	.39

Not: Ss= Standart sapma, G= ayırma oranı, Sd= serbestlik derecesi

Tablo 9 incelendiğinde sporculara ilişkin ayırma indeksi güvenilirliğinin tüm yıllarda ve gruplarda 1'e eşit olduğu görülmektedir. Ayırma oranı (G) ise 14.60 ile 37.75 arasında değişmekte ve oldukça yüksek değerler almaktadır. Yani bir sporcunun yeteneği diğer sporcunun yeteneğinden ayrılmaktadır. Bu da merkezi eğilim hatasının sorun olmadığını kanıtlamaktadır (Myford ve Wolfe, 2004). Bu sonuç anlamlı tamamı aynı (sabit etki) X^2 değeri ile desteklenmektedir. Tamamı aynı ki-kare değeri istatistiksel olarak anlamlı olduğunda bireylerin yetenek düzeylerinin birbirlerinden başarılı bir biçimde ayrıldığını göstermektedir. Tablo 9'da bireylere ilişkin sonuçlar incelendiğinde tüm yıl ve gruplarda tamamı aynı ki-kare değeri anlamlı bulunmuştur. Rastgele normal ki-kare ise sporcuların normal dağılım gösterip göstermediğini ve seçkisiz örneklem olup olmadığını test etmektedir (Myford ve Wolfe, 2003). Rastgele normal ki-kare değerinin istatistiksel olarak anlamlı bulunmaması, sporcuların normal dağılıma sahip olduğunu ve model veri uyumunun sağlandığını göstermektedir.

Ancak diğer değişkenlik kaynakları söz konusu olduğunda ayırma indeksi güvenilirliği ve ayırma oranındaki düşük değerler yüksek düzeyde tutarlılığı göstermektedir. Tablo 10'da ise görevler için ayırma indeksi oranı, ayırma indeksi güvenilirliği, sabit etki ve rastgele etki hipotezine ilişkin X^2 testi sonuçları yer almaktadır.

Tablo 10. 2006–2008 Dünya Buz Pateni Şampiyonası serbest programa ilişkin görevler için ayırma oranı, güvenilirlik ve X^2 testleri.

	RMSEA	Ss	G	Güvenirlik	Tamamı aynı X^2	Sd	p	Rastgele Normal X^2	Sd	p
2006 Erkek	.05	.17	3.61	.93	70.6	4	.00	3.8	3	.28
2006 Kadın	.05	.29	6.31	.98	203.9	4	.00	3.9	3	.27
2006 Çift	.05	.31	5.74	.97	169.7	4	.00	3.9	3	.27
2007 Erkek	.05	.36	7.72	.98	306.7	4	.00	3.9	3	.27
2007 Kadın	.04	.27	6.92	.98	250.0	4	.00	3.9	3	.27
2007 Çift	.05	.25	4.78	.96	119.8	4	.00	3.9	3	.27
2008 Erkek	.04	.27	6.93	.98	250.2	4	.00	3.9	3	.27
2008 Kadın	.04	.29	7.06	.98	260.5	4	.00	3.9	3	.27
2008 Çift	.04	.22	4.91	.96	126.0	4	.00	3.9	3	.27

Not: Ss= Standart sapma, G= ayırma oranı, Sd= serbestlik derecesi

Tablo 10 incelendiğinde, görevlere ilişkin ayırma indeksi güvenilirliği .93 ile .98 arasında, ayırma oranı ise 3.61 ile 7.72 arasında değişmektedir. Bu da görevler arasında anlamlı fark olduğu ve birinin diğeri yerine kullanılmayacağını göstermektedir. Bu sonuç anlamlı tamamı aynı ki-kare değeri ile de desteklenmektedir. Böylelikle, bazı görevler diğelerinden daha yüksek yetenek gerektirmektedir. Anlamlı olmayan rastgele normal ki-kare değerleri görevlerin normal dağılıma sahip olduğu şeklinde yorumlanabilir.

Tablo 11’de hakemler için ayırma indeksi oranı, ayırma indeksi güvenilirliği, sabit etki ve rastgele etki hipotezine ilişkin X^2 testi sonuçları yer almaktadır.

Tablo 11. 2006–2008 Dünya Buz Pateni Şampiyonası serbest programa ilişkin hakemler için ayırma oranı, güvenilirlik ve X^2 testleri.

	RMSEA	Ss	G	Güvenirlik	Tamamı aynı X^2	Sd	p	Rastgele Normal X^2	Sd	p
2006 Erkek	.07	.18	2.41	.85	82.0	11	.00	9.7	10	.47
2006 Kadın	.07	.30	4.21	.95	225.7	11	.00	10.5	10	.40
2006 Çift	.08	.28	3.38	.92	147.8	11	.00	10.2	10	.42
2007 Erkek	.07	.26	3.60	.93	167.0	11	.00	10.3	10	.41
2007 Kadın	.06	.18	2.93	.90	116.4	11	.00	10.1	10	.44
2007 Çift	.08	.13	1.58	.71	41.8	11	.00	8.7	10	.56
2008 Erkek	.06	.27	4.39	.95	251.5	11	.00	10.5	10	.39
2008 Kadın	.06	.40	6.18	.97	472.5	11	.00	10.8	10	.38
2008 Çift	.07	.38	5.49	.97	380.7	11	.00	10.7	10	.38

Not: Ss= Standart sapma, G= ayırma oranı, Sd= serbestlik derecesi

Tablo 11 incelendiğinde, hakemler için ayırma indeksi güvenilirliği .71 ile .97 arasında ayırma oranı ise 1.58 ile 6.18 arasında değişmektedir. Sadece 2007 çiftler serbest programda ayırma indeksi güvenilirliği (.71, $G=1.58$) diğer yıllara göre daha düşüktür. McNamara (1996)'ya göre, güvenilirlik indeksinin geleneksel anlamından farklı olduğunu, yani hakemler arası uyum düzeyini vermediğini, hakemlerin katılık düzeylerinin gerçek farklılığını verdiğini söylemektedir. Bu sporcuları genel sıralamalarındaki farklılığı göstermemektedir. Bu değer, hakemlerin, birbirilerine ne kadar benzedikleri yerine birbirlerinden ne kadar farklı olduğunu vermektedir. Bu değer, hakemler arası güvenilirlik veya uyum olarak düşünülmemelidir. Ayırma güvenilirliğinin 0'a yakın olması, hakemlerin birbiri yerine kullanılabilceğini, katılık/cömertlik açısından fark olmadığını göstermektedir (Myford ve Wolfe, 2003; Sudweeks, Reeve ve Bradshaw, 2005). Bu yüksek değerler, hakemler arasında gerçek farklılıkların olduğunu belirtmektedir. Bunun anlamı, hakemlerin katılık/cömertlikleri arasında yüksek düzeyde istenmeyen varyans olduğudur (Engelhard, 2002; Hoyt, 2000). Hakemlere ilişkin tamamı aynı ki-kare değerleri ile rastgele normal ki-kare değerleri incelendiğinde; tamamı aynı ki-kare değerlerinin anlamlı, rastgele normal ki-kare değerlerinin ise anlamlı olmadığı görülmektedir. Buna göre; hakemlerin katılık/cömertlikleri arasında anlamlı fark olmasına rağmen, hakemler normal dağılım göstermektedirler.

Alt Problem IV: 2009- 2011 yılları arasında buz pateni performansının ölçülmesinde 9 hakemin kullanıldığı yarışmalar (model veri uygunluğu, sporcuların yetenekleri ve uygunluk istatistikleri, hakemlerin katılık/cömertlik istatistikleri, hakemxsporcu yanlılığı) için ÇDKRM analizi sonuçları:

Dünya Buz Pateni Şampiyonalarında, 2009–2011 yıllarında erkekler, kadınlar ve çiftler serbest program yarışmalarında elde edilen sonuçlara ilişkin verilerin modelle uyumunu test etmek amacı ile ÇDKRM analizleri sonucunda elde edilen standartlaştırılmış artık değerleri incelenmiştir. Buna ilişkin sonuçlar Tablo 12'de verilmiştir.

Tablo 12. 2009–2011 Dünya Buz Pateni Şampiyonaları serbest program değerlendirmelerine ilişkin ÇDKRM analizleri sonucunda elde edilen standart artık yüzdeleri, sporcu, görev ve hakem değişkenlik kaynaklarına göre uyumsuzluk yüzdeleri

Yıl	Bireyler	Standart Artık (%)	Uyumsuzluk Sporcular	Uyumsuzluk Görevler	Uyumsuzluk Hakemler
2009	Erkek	7 (%0.6)	n=6 (%25) 1,2,5,10,13,21	—	n=1 2
	Kadın	10 (%0.9)	n=7 (%29.2) 2,3,8,9,14,15,22	—	n=4 1,4,7,8
	Çift	4 (%0.4)	n=5 (%25) 4,8,13,14,16	—	n=3 3,8,9
2010	Erkek	13 (%1.2)	n=8 (%33.3) 4,6,8,9,13,17,22,24	n=1 geçişler	n=2 5,6
	Kadın	4 (%0.4)	n=7 (%29.2) 1,6,9,14,20,21,23	—	n=1 6
	Çift	1 (%0.1)	n=3 (%18.8) 5,8,14	—	n=1 5
2011	Erkek	5 (%0.5)	n=5 (%20.8) 4,6,9,17,20	—	n=1 2
	Kadın	3 (%0.3)	n=5 (%20.8) 4,5,14,18,22	—	n=1 3
	Çift	4 (%0.6)	n=4 (%25) 2,5,7,13	—	n=2 4,7

Tablo 12 incelendiğinde ± 3 aralığının dışında olan standartlaştırılmış artık değerlerinin oranının %0.1 ile %1.2 arasında değiştiği bulunmuştur. Sadece 2010 erkekler serbest programda elde edilen verinin %1.2 olduğu görülmektedir. Ancak çok aşırı bir sapma olmadığından dolayı bu verilerin modele uygun olduğu kabul edilebilir (Looney, 2012). Ayrıca, Tablo 12’de sporcular, görevler ve hakemler için uyumsuzluk sayıları ve uyumsuzluk gösteren sporcular, görevler ve hakemler belirtilmiştir. Bu sayılar belirlenirken iç uyum ve dış uyum kareler ortalamasının 0.40’dan küçük veya 1.2’den büyük olduğu durumlar dikkate alınmıştır. Görevler açısından incelendiğinde 2010 erkekler serbest programda tek bir görev açısından uyumsuzluk belirlenmiştir. Bu uyumsuzluğun, geçişlerde olduğu tespit edilmiştir. Bu bulgu da, Looney (2012)’in Olimpiyat Buz Pateni verileri ile yaptığı çalışması sonucunda elde ettiği bulgular tarafından desteklenmektedir.

Hakemler açısından incelendiğinde ise; uyumsuzluk gösteren hakem sayısının 1 ile 4 arasında değiştiği görülmektedir. 2006–2008 yıllarına göre uyum değerleri aralığı dışında kalan hakem sayısının daha az olduğu görülmektedir. Tablo 10, hakemlerin tüm sporcuları ve tüm program bileşenlerini puanlamadaki tutarlılığına ilişkin bilgi sağladığından, tabloda belirtilen hakemlerin kendi içlerindeki puanlamalarının tutarlı olmadığı söylenebilir. Ancak uyum istatistikleri, hakemlerin tüm sporcuları puanlamada tutarlı olup olmadığını belirlemede yeterli olmadığından yanlılık analizi de yapılmıştır. Bunun için sporcuhakem yanlılık oranları hesaplanmıştır. Buna göre; 2009 erkekler serbest programda %11.5, 2009 kadınlar serbest programda %6.1, 2009 çiftler serbest programda %5.0; 2010 erkekler serbest programda %11.1, 2010 kadınlar serbest programda %13.7, 2010 çiftler serbest programda %5.3; 2011 erkekler serbest programda %11.2, 2011 kadınlar serbest programda %6.9, 2011 çiftler serbest programda %6.9 düzeyinde hakemler sporcuları farklı puanlamışlardır. Bu oranlar 2006–2008 yıllarında elde edilen oranlara göre daha yüksektir.

Sporcular açısından incelendiğinde ise; uyumsuzluk gösteren sporcu sayısının 3 ile 8 arasında değiştiği görülmektedir. Bu da, bu sporcuların performansındaki değişkenliğin beklenen modelden anlamlı olarak farklı olduğunu göstermektedir. Sporculara ilişkin uyum istatistiklerinin istenilen sınırlarda olmaması o sporcuya ilişkin yetenek kestirimine güvenilmeyeceği anlamına gelmektedir. Ancak, sporculara ilişkin uyum istatistikleri incelendiğinde; uyum istatistikleri 2'nin üzerinde olan çok fazla sporcu olmadığından dolayı, bu sonuçların genel sonuçları bozmadığı düşünülmektedir (Sudweeks, Reeve ve Bradshaw, 2005).

Tablo 13'te 2009–2011 yılları arasında erkekler, kadınlar ve çiftler serbest program yarışmalarına ilişkin ÇDKRM analizleri sonucunda elde edilen en düşük ve en yüksek logit değerleri, logit ortalama ve standart sapmaları, iç uyum ve dış uyum kareler ortalamasına ait ortalama ve standart sapma değerleri verilmektedir. Tabloda her bir değişkenlik kaynağı için bu değerler ayrı ayrı verilmiştir (Sporcuların yeteneklerine, görevlerin güçlüğüne ve hakemlerin katılığına ilişkin bireysel düzeydeki logit, iç uyum ve dış uyum değerleri EK-2'de verilmiştir).

Tablo 13. 2009–2011 Dünya Buz Pateni Şampiyonaları serbest program değerlendirmelerine ilişkin ÇDKRM değişkenlik kaynaklarına ait logit minimum/maksimum, logit, iç uyum ve dış uyum ortalama ve standart sapma değerleri

		Logit min-max	Logit Ort.± Ss	İç uyum Ort± Ss	Dış uyum Ort ±Ss
2009 Erkek	Sporcular	(-2.62)-(-2.27)	-.18±1.48	.99±.48	.99±.48
	Görevler	(-.46)-(.30)	.00±.25	.98±.11	.99±.11
	Hakemler	(-.42)-(.35)	.00±.20	.98±.63	.99±.66
2009 Kadın	Sporcular	(-2.56)-(-4.10)	.33±1.81	.99±.53	.99±.52
	Görevler	(-.42)-(.22)	.00±.22	.97±.08	.99±.10
	Hakemler	(-.60)-(.43)	.00±.38	.97±.56	.99±.55
2009 Çift	Sporcular	(-4.32)-(-4.45)	-.59±2.33	.98±.25	.97±.25
	Görevler	(-.33)-(.25)	.00±.19	.98±.11	.97±.11
	Hakemler	(-.38)-(.47)	.00±.27	.98±.35	.97±.34
2010 Erkek	Sporcular	(-2.05)-(-2.93)	-.01±1.19	.99±.33	.98±.32
	Görevler	(-.35)-(.19)	.00±.18	1.00±.20	.98±.21
	Hakemler	(-.06)-(.10)	.00±.05	1.00±.27	.98±.25
2010 Kadın	Sporcular	(-1.73)-(-2.16)	.16±1.07	.99±.30	.99±.30
	Görevler	(-.30)-(.22)	.00±.17	.99±.15	.99±.14
	Hakemler	(-.14)-(.14)	.00±.09	.99±.12	.99±.10
2010 Çift	Sporcular	(-3.24)-(-4.48)	-.03±2.65	.96±.20	.96±.20
	Görevler	(-.39)-(.31)	.00±.23	.96±.09	.96±.09
	Hakemler	(-.23)-(.38)	.00±.18	.96±.23	.96±.23
2011 Erkek	Sporcular	(-2.05)-(-2.50)	-.14±1.13	.99±.38	.99±.39
	Görevler	(-.42)-(.19)	.00±.22	.98±.14	.99±.13
	Hakemler	(-.16)-(.24)	.00±.13	.98±.14	.99±.14
2011 Kadın	Sporcular	(-3.84)-(-4.50)	.12±2.39	.98±.30	.98±.30
	Görevler	(-.50)-(.24)	.00±.26	.99±.11	.98±.10
	Hakemler	(-.21)-(.22)	.00±.12	.99±.20	.98±.20
2011 Çift	Sporcular	(-2.74)-(-3.95)	.02±2.04	.98±.29	.98±.28
	Görevler	(-.33)-(.19)	.00±.18	.98±.09	.98±.09
	Hakemler	(-.16)-(.16)	.00±.10	.98±.34	.98±.31

Not: Ortalamanın +1 ve -1 standart sapma aralığı verilmiştir.

Tablo 13 incelendiğinde, iç uyum kareler ortalaması ve dış uyum kareler ortalamasına ilişkin ortalamaların tüm gruplarda ve tüm değişkenlik kaynaklarında 1'e yakın olduğu görülmektedir. Bu da verilerin model ile uyumlarının iyi olduğuna işaret etmektedir. Yani başka bir deyişle, veriler ölçme için uygundur (Linacre, 2007).

Tablo 14'te ise sporcular için ayırma indeksi oranı, ayırma indeksi güvenilirliği, sabit etki ve rastgele etki hipotezine ilişkin X^2 testi sonuçları yer almaktadır.

Tablo 14. 2009–2011 Dünya Buz Pateni Şampiyonası serbest programa ilişkin sporcular için ayırma oranı, güvenilirlik ve X^2 testleri.

	RMSEA	Ss	G	Güvenirlik	Tamamı aynı X^2	Sd	p	Rastgele Normal X^2	Sd	p
2009 Erkek	.10	1.47	15.36	1.00	5995.2	23	.00	22.9	22	.41
2009 Kadın	.09	1.80	19.65	1.00	9034.1	23	.00	22.9	22	.41
2009 Çift	.10	2.32	22.13	1.00	8791.4	19	.00	19.0	18	.39
2010 Erkek	.08	1.19	14.52	1.00	4115.2	23	.00	22.9	22	.41
2010 Kadın	.07	1.07	14.23	1.00	5115.1	23	.00	22.9	22	.41
2010 Çift	.11	2.65	25.01	1.00	9678.4	15	.00	15.0	14	.38
2011 Erkek	.08	1.12	13.51	.99	4224.6	23	.00	22.9	22	.41
2011 Kadın	.11	2.39	21.61	1.00	11200.8	23	.00	23.0	22	.40
2011 Çift	.09	2.03	22.41	1.00	7304.9	15	.00	15.0	14	.38

Not: Ss= Standart sapma, G= ayırma oranı, Sd= serbestlik derecesi

Tablo 14 incelendiğinde sporculara ilişkin ayırma indeksi güvenirliliğinin tüm yıllarda ve gruplarda .99 ile 1 arasında değiştiği görülmektedir. Ayırma oranı (G) ise 13.51 ile 25.01 arasında değişmekte ve oldukça yüksek değerler almaktadır. Yani bir sporcunun yeteneği diğer sporcunun yeteneğinden ayrılmaktadır. Bu da merkezi eğilim hatasının sorun olmadığını kanıtlamaktadır (Myford ve Wolfe, 2004). Ancak bu değerlerin 2006–2008 yıllarında ait verilerde elde edilen sonuçlardan daha düşük olduğunu gözlemleyiz. Bunun nedeni de hakem sayısındaki değişim (12’den 9’a indirilmesi) olabilir. Bu sonuç anlamlı tamamı aynı (sabit etki) X^2 değeri ile desteklenmektedir. Tamamı aynı ki-kare değeri istatistiksel olarak anlamlı olduğunda bireylerin yetenek düzeylerinin birbirlerinden başarılı bir biçimde ayrıldığını göstermektedir. Tablo 12’de bireylere ilişkin sonuçlar incelendiğinde tüm yıl ve gruplarda tamamı aynı ki-kare değeri anlamlı bulunmuştur. Rastgele normal ki-kare ise sporcuların normal dağılım gösterip göstermediğini ve seçkisiz örneklem olup olmadığını test etmektedir (Myford ve Wolfe, 2003). Anlamlı bulunmaması, sporcuların normal dağılıma sahip olduğunu ve model veri uyumunun sağlandığını göstermektedir.

Tablo 15’te ise görevler için ayırma indeksi oranı, ayırma indeksi güvenirliliği, sabit etki ve rastgele etki hipotezine ilişkin X^2 testi sonuçları yer almaktadır.

Tablo 15. 2009–2011 Dünya Buz Pateni Şampiyonası serbest programa ilişkin görevler için ayırma oranı, güvenilirlik ve X^2 testleri.

	RMSEA	Ss	G	Güvenirlik	Tamamı aynı X^2	Sd	p	Rastgele Normal X^2	Sd	p
2009 Erkek	.04	.24	5.60	.97	164.0	4	.00	3.9	3	.27
2009 Kadın	.04	.22	5.32	.97	147.5	4	.00	3.9	3	.27
2009 Çift	.05	.18	3.51	.93	66.9	4	.00	3.8	3	.29
2010 Erkek	.04	.18	4.96	.96	131.4	4	.00	3.9	3	.27
2010 Kadın	.03	.17	4.87	.96	125.1	4	.00	3.9	3	.27
2010 Çift	.06	.22	3.87	.94	83.5	4	.00	3.8	3	.28
2011 Erkek	.04	.21	5.60	.97	160.5	4	.00	3.9	3	.27
2011 Kadın	.05	.25	5.10	.96	135.1	4	.00	3.9	3	.27
2011 Çift	.05	.17	3.50	.92	67.9	4	.00	3.8	3	.29

Not: Ss= Standart sapma, G= ayırma oranı, Sd= serbestlik derecesi

Görevler açısından ayırma indeksi güvenilirliği ve ayırma oranı değerleri incelendiğinde (Tablo 15); görevlere ilişkin ayırma indeksi güvenilirliği .92 ile .97 arasında, ayırma oranı ise 3.50 ile 5.60 arasında değişmektedir. Yüksek ayırma indeksi güvenilirliği ve ayırma oranı değerleri görevler arasında anlamlı fark olduğu ve birinin diğeri yerine kullanılmayacağını göstermekteyken, düşük değerler ise görevlerin birbirine benzediğini ve birbirileri yerine kullanılacaklarını göstermektedir. Bir diğeri ifadeyle, yüksek değerler; bazı görevlerin daha fazla yetenek gerektirdiği şeklinde yorumlanabilirken, düşük değerler ise tüm görevlerin benzer düzeyde yetenek gerektirdiği şeklinde yorumlanabilir. Görevlerin, sporcuları eşit düzeyde zorladığı anlamına gelmektedir. Bu sonuçlar, tamamı aynı ki-kare değerleri ile de desteklenmektedir. Anlamlı olmayan rastgele normal ki-kare değeri görevlerin normal dağılıma sahip olduğu şeklinde yorumlanabilir.

Tablo 16'da hakemler için ayırma indeksi oranı, ayırma indeksi güvenilirliği, sabit etki ve rastgele etki hipotezine ilişkin X^2 testi sonuçları yer almaktadır.

Tablo 16. 2009–2011 Dünya Buz Pateni Şampiyonası serbest programa ilişkin hakemler için ayırma oranı, güvenilirlik ve X^2 testleri.

	RMSEA	Ss	G	Güvenirlik	Tamamı aynı X^2	Sd	p	Rastgele Normal X^2	Sd	p
2009 Erkek	.06	.20	3.35	.92	110.9	8	.00	7.5	7	.38
2009 Kadın	.06	.38	6.86	.98	434.9	8	.00	7.9	7	.35
2009 Çift	.07	.26	3.76	.93	136.7	8	.00	7.6	7	.37
2010 Erkek	.05	.01	.15	.02	9.2	8	.33	4.3	7	.75
2010 Kadın	.05	.08	1.79	.76	37.8	8	.00	6.6	7	.47
2010 Çift	.08	.17	2.14	.82	49.2	8	.00	6.9	7	.44
2011 Erkek	.05	.12	2.26	.84	55.2	8	.00	7.0	7	.43
2011 Kadın	.07	.10	1.50	.69	29.1	8	.00	6.3	7	.51
2011 Çift	.07	.07	1.12	.56	20.5	8	.00	5.8	7	.57

Not: Ss= Standart sapma, G= ayırma oranı, Sd= serbestlik derecesi

Tablo 16 incelendiğinde, hakemler için ayırma indeksi güvenilirliği .02 ile .98 arasında ayırma oranı ise .15 ile 6.86 arasında değişmektedir. Ayırma güvenirliliğinin 0'a yakın olması, hakemlerin birbiri yerine kullanılabileceğini, katılık/cömertlik açısından fark olmadığını göstermektedir (Myford ve Wolfe, 2003; Sudweeks, Reeve ve Bradshaw, 2005). 2010 erkekler serbest programa ilişkin değerler incelendiğinde ayırma indeksi güvenirliliğinin .02, ayırma oranının da .15 olduğu görülmektedir. Ayrıca tamamı aynı ki-kare değeri de anlamlı bulunmamıştır (Tamamı aynı X^2 : 9.2 sd: 8 p=.33). Bu düşük değerler, hakemler söz konusu olduğunda arzu edilen bir durumdur. Yüksek değerler, hakemler arasında gerçek farklılıkların olduğunu belirtmektedir. Bunun anlamı, hakemlerin katılık/cömertlikleri arasında yüksek düzeyde istenmeyen varyans olduğudur (Engelhard, 2002; Hoyt, 2000). Hakemlere ilişkin tamamı aynı ki-kare değerleri ile rastgele normal ki-kare değerleri incelendiğinde; tamamı aynı ki-kare değerlerinin anlamlı, rastgele normal ki-kare değerlerinin ise anlamlı olmadığı görülmektedir. Buna göre; hakemlerin katılık/cömertlikleri arasında anlamlı fark olmasına rağmen, hakemler normal dağılım göstermektedirler.

Alt Problem V: Genellenebilirlik Kuramı ve Çok Değişkenlik Kaynaklı Rasch Modeline göre Sporcu ve Hakem Değişkenlik Kaynaklarında Elde Edilen Parametrelerin Karşılaştırılması

Dünya Buz Pateni Şampiyonalarından 2006–2011 yılları arasında erkekler, kadınlar ve çiftler serbest programda sporcuların performansının ölçülmesine ilişkin olarak yapılan genellenebilirlik kuramı ve çok değişkenlik kaynaklı Rasch modeli analizleri sonucunda elde edilen istatistikler bu bölümde karşılaştırılmıştır.

Genellenebilirlik Kuramı ve ÇDKRM'ni karşılaştırmada öncelikle her ikisinin de kullandığı kavramları açıklamakta fayda vardır. Bunlar; değişkenlik kaynağı, etkileşim ve güvenirliliktir.

Genellenebilirlik kuramı değişkenlik kaynaklarını sistematik hata varyansının bir kaynağı olarak tanımlamaktadır. Ölçme nesnesi tarafından açıklanan varyans arzu edilen bir durumdur. Değişkenlik kaynakları da modele katkısı bulunan diğer

bileşenlerdir (ölçme nesnesi dışında kalanlar) ve bunlar da puan değişkenliğine katkıda bulunan dış varyanslardır. ÇDKRM’nde ise değişkenlik kaynakları modeldeki sistematik değişkenliğin kaynakları olarak tanımlanmaktadır. Örneğin; sporcu, görev ve hakemden oluşan bir model kurduğumuzda, G-kuramı sporcuları ölçme nesnesi olarak, görev ve hakemleri ise değişkenlik kaynağı olarak ele almaktadır. Ancak ÇDKRM’nde üçü de değişkenlik kaynağı olarak düşünülmektedir (Sudweeks, Reeve ve Bradshaw, 2005).

Etkileşim kavramı hem G-kuramında hem de ÇDKRM’nde birbirilerinden biraz farklı şekillerde kullanılmaktadır. ÇDKRM’nde değişkenlik kaynakları arasındaki etkileşimler bireyin farklı fonksiyonlaşmasını, maddenin farklı fonksiyonlaşmasını vb temsil etmektedir. Her biri durumdan duruma tek tek rapor edilmektedir (Linacre, 1989). G kuramında ise etkileşimler varyansın faktöryel analizi şeklinde tanımlanmaktadır. Bu etkileşim varyanslarının karşılaştırma büyüklükleri bir değişkenlik kaynağındaki görelî sıranın, diğer değişkenlik kaynağındaki farklılığını belirtmektedir (Shavelson ve Webb, 1991). Örneğin; hakemlere ilişkin ana etki, örneklemdaki hakemlerin katılık düzeylerindeki farklılığı temsil ederken, sporcu hakem etkileşimi, sporcunun farklı hakemler tarafından farklı puanlandığını göstermektedir.

Güvenirlilik kavramı açısından incelendiğinde, genellenebilirlik kuramı G ve phi katsayılarını hesaplamaktadır. Her ikisi de güvenirlilik katsayısı olarak yorumlanabilmektedir. Bu katsayılar bireyin puanlama ortalamasının, bireyin evren puanını ne kadar iyi yordadığına ilişkin bilgi vermektedir. Tek değişkenlik kaynağının olduğu modelde, g katsayısı Cronbach alfaya eşittir. Değişkenlik kaynaklarının sayısı arttıkça g-katsayısı daha küçük değerler almaktadır. Phi-katsayısı da daha fazla hata varyansı içerdiğinden g- katsayısına göre daha küçük çıkmaktadır.

ÇDKRM’nde ise iki farklı güvenirlilik katsayısı bulunmaktadır. Bunlar; ayırma indeksinin güvenirliliği ve ayırma oranıdır. Bu iki istatistik farklı ölçeklerde rapor edilmektedir. Aynı bilgiler kullanılarak hesaplanmakta ve benzer sonuçlara ulaşmaktadır. ÇDKRM ile kestirilen ölçümlerin değişkenliğine ilişkin bilgi vermektedirler. Ayırma indeksinin güvenirliliği 0 ile 1.0 arasında değer alırken, ayırma oranı ise 0 ile sonsuz arasında değer almaktadır. Her bir değişkenlik kaynağı için bu değerler hesaplanmaktadır. Ancak bu

değerlerin yorumları her bir değişkenlik kaynağı için farklılaşmaktadır. Birey değişkenlik kaynağı söz konusu olduğunda; ayırma indeksinin güvenilirliği ve ayırma oranının yüksek olması arzu edilen bir durumken, diğer değişkenlik kaynaklarında (hakemler) düşük olması arzu edilen bir durumdur (Myford ve Wolfe, 2003; Sudweeks, Reeve ve Bradshaw, 2005; Bond ve Fox, 2007).

Elde edilen bulgular incelendiğinde, - kuramı ve ÇDKRM analizlerinin ikisi de toplam varyansı açıklamada sporcu değişkenlik kaynağının sahip olduğu payın yüksek olduğunu belirlemişlerdir. Genellenebilirlik analizleri incelendiğinde, sporcu ana etkisi yıllara ve gruplara göre %77.6 ile %94.4 arasında değişmekte ve tüm genellenebilirlik analizlerinde ilk sırada yer almaktadır. Sporcu, yapılan çalışmada ölçme nesnesi olduğundan dolayı, bu istenilen bir durumdur. Sporcular arası farklılıkların belirlenebildiğini göstermektedir. Genellenebilirlik analizleri sonucunda, g-katsayısı (görelî) ve phi katsayısı (mutlak) da elde edilmektedir. Bu çalışmada, G (.98-1.00) ve phi katsayıları (.97-1.00) oldukça yüksek bulunmuştur. Bu değerler, buz pateni performansına ilişkin güvenilir sonuçlar elde edildiğine dair bilgi vermektedir. ÇDKRM’nde ise sporcu değişkenlik kaynağına ilişkin istatistikler incelendiğinde, ayırma indeksi güvenilirliğinin tüm yıllarda ve gruplarda .99 ile 1.00 arasında değiştiği görülmektedir. Ayırma oranı (G) ise 13.51 ile 37.75 arasında değişmektedir. Bu değerler, sporcuların buz patenine ilişkin yetenekleri açısından birbirinden ayrıldığını göstermektedir. Elde edilen bu sonuç, tamamı aynı ki-kare testi ile de desteklenmiştir. Tamamı aynı ki-kare testi, tüm analizlerde anlamlı bulunmuştur. İstatistiksel olarak anlamlı olması, bireylerin yetenek düzeylerinin birbirlerinden başarılı bir biçimde ayrıldığını göstermektedir. Bu bulgu, genellenebilirlik kuramı bulguları ile uyum göstermektedir.

Araştırmada incelenen bir diğer değişkenlik kaynağı ise hakemlerdir. Genellenebilirlik çalışmaları sonucunda, hakem ana etkisinin, toplam varyansı açıklama oranı 2006–2011 yıllarındaki tekler ve çiftler serbest program yarışmalarında %0.0 ile %3.7 arasında bulunmuştur. Varyansın düşük olması, hakemler arasındaki değişkenliğin az olduğunu göstermektedir. Buna göre, hakemler sporcuları puanlamada tutarlıdırlar. ÇDKRM’nde ise uyum istatistikleri hakemlerin sporcuları puanlamadaki tutarlılıklarını vermektedir.

Yani tüm sporcuları puanlamada, hakemin kendi içindeki tutarlılığına ilişkin bilgi vermektedir. Hakemlere ilişkin uyum istatistikleri incelendiğinde, iç uyum ve dış uyum kareler ortalamasına ait ortalama değerleri 1'e yakın bulunmuştur (.90 ile 1.0 arasında). Bu da hakemlerin puanlama yaparken, zor göreve düşük, kolay göreve yüksek puan; yetenekli sporcuya yüksek, daha az yetenekliye düşük puan verdikleri anlamına gelmektedir. ÇDKRM'nde elde edilen diğer istatistikler incelendiğinde, genellikle hakemlere ilişkin ayırma indeksi güvenilirliklerinin ve ayırma oranlarının yüksek olduğu bulunmuştur. Anlamlı sabit etki (tamamı aynı) ki-kare değeri de bu bulguları desteklemektedir. Ayırma indeksi güvenirliliği ve ayırma oranı hakemlerin katılık/cömertlik açısından birbirlerinden farklı olduğunu vermektedir. Hakemler birbirileri yerine kullanılamayacağı şeklinde yorumlanmaktadır. Sadece 2010 erkekler serbest programda, ayırma indeksi güvenirliliği ve ayırma oranı arzu edildiği gibi düşük bulunmuş ve tamamı aynı ki-kare değeri de istatistiksel olarak anlamlı bulunmamıştır.

Genellenebilirlik kuramı analizlerine ait bulgular incelendiğinde, sporcuxhakem etkileşiminin, toplam varyansı açıklamada ikinci sırada yer aldığını görmekteyiz. Bu etkileşime ilişkin varyans, erkekler serbest programda %8.8-%12.4, kadınlar serbest programda %6.8-%14.4 ve çiftler serbest programda %2.8-%7.4 arasındadır. Sporcuxhakem etkileşimi, belirli bir sporcuyu puanlamada, bazı hakemlerin katı, bazı hakemlerin ise cömert davrandığını göstermektedir. ÇDKRM'nde elde edilen hakem istatistikleri, G-kuramına ilişkin analizler sonucunda ulaşılan bu sonuca ilişkin bilgi vermekte yeterli olmadığından dolayı (Lynch ve McNamara, 1998), sporcuxhakem etkileşimine ilişkin yanlılık analizi yapılmıştır. Bu analiz sonucuna göre, sporcuxhakem etkileşimine ait yanlılık oranları; erkekler serbest programda %8.5-%11.5, kadınlar serbest programda %6.1-%13.7 ve çiftler serbest programda %2.8- %6.9'dur. Elde edilen bu bulguların genellenebilirlik analizleri ile benzer nitelikte olduğu söylenebilir.

G-kuramı ile ÇDKRM analizleri hangi ögenin değişkenliğe katkısının en yüksek olduğunu belirlemede aynı sonuca ulaştıkları görülmektedir (Smith ve Kulikowich, 2004). Diğer çalışmaların da ortaya koyduğu gibi iki yaklaşımın da farklı avantajlarının olduğu gözlenmektedir (Alharby, 2006; Lynch ve McNamara, 1998; Smith ve Kulikowich, 2004; Sebok, 2010; Sudweeks, Reeve ve Bradshaw, 2005; Ure, 2011).

G-kuramı her bir deęişkenlik kaynaęı için ana etki ve etkileşim varyansı vermektedir. Ayrıca G ve phi katsayıları hesaplanmaktadır. Bu da mutlak ve görelî kararlar verilirken genellenebilirlik düzeyini göstermektedir. Ayrıca karar çalıřmaları kullanarak, sporcu, hakem ve görev sayılarında deęişiklikler yapıldığında güvenilirlikteki deęişimleri yansıtmaktadır.

ÇDKRM analizleri bireysel düzeydedir. Elde edilen çıktıları, her bir deęişkenlik kaynaęını örneklemeden bağımsız ve eşit aralıklı logit ölçeğinde raporlaştırmaktadır. ÇDKRM ayrıca her bir deęişkenlik kaynaęına ilişkin öğelerin olası puan doğrusunda nasıl dağıldığını gösteren ayırma indeksi deęerini vermektedir. Ayrıca model veri uyumuna ilişkin (iç uyum ve dıř uyum istatistikleri) bilgi vermektedir.

G- kuramı ve ÇDKRM deęişkenlik kaynaklarının görelî büyüklükleri konusunda uyum göstermektedirler. Bu teknikler, deęişkenlik kaynaklarını farklı şekillerde ele almaktadırlar. G-kuramı, çeşitli görevlerin ve farklı hakemlerin bir biri yerine kullanılmaları amacı ile görev ve hakemlerin homojenliğini vurgulamaktadır. ÇDKRM ise görevlerin heterojenlięi ile hakemlerin kendi içerisindeki tutarlılıęına vurgu yapmaktadır. Ayrıca, sporcular grubunun gözlenen ham puanları ile bu gruba benzer sporculardan oluşan ve benzer koşullar altında elde edilen ham puanlara sahip bir grup ile arasındaki benzerliklerin kestirimi önemli olduęu durumlarda G-kuramı kullanılabilir. Ancak ÇDKRM'nin G-kuramına göre avantajları bulunmaktadır. Bunlar; hakem ve görev kalibrasyonları performans deęerlendirmelerindeki eşitlięe yardımcı olmaktadır, yapılan dönüşümler sonucunda sıralı ham puanlar, eşit aralıklı puanlara dönüřtürölmektedir, ölçümler farklı durumlara genellenebilmektedir, grup performansı yerine her bir deęişkenlik kaynaęındaki her bir öęe için bilgi sağlamaktadır. Tüm bunlar dikkate alındığında, G-kuramı ve ÇDKRM'nin bir arada kullanımını uygundur.

BÖLÜM IV

SONUÇ ve ÖNERİLER

Bu bölümde, araştırmada elde edilen bulgulara ilişkin sonuçlara ve sonuçlara bağlı olarak önerilere yer verilmiştir.

4.1 SONUÇLAR

I. 2006–2011 yıllarında yapılan Dünya Buz Pateni Şampiyonalarında, tekler (kadınlar ve erkekler) ve çiftler serbest programda, buz pateni performansının ölçülmesinde yer alan 5 göreve (kayma becerileri, geçişler, yorumlama, koreografi ve performans) tamamen çaprazlanmış s x g x h modeli uygulanarak yapılan toplam 18 genellenebilirlik çalışması sonucunda, tüm yıllarda ve tüm gruplarda sporcu (s) ana etkisi için kestirilen varyans bileşeninin toplam varyansın büyük bir oranını açıkladığı ve toplam varyans içerisinde ilk sırada yer aldığı görülmüştür. Bu araştırmada, sporcular için kestirilen varyansın toplam varyans içindeki oranının büyük olması nedeniyle buz pateni performansının ölçülmesinde yer alan 5 görevde, bireyler arası farklılıkların ortaya çıkarılabildiği sonucuna varılmıştır. Görev ana etkisi için kestirilen varyans, toplam varyansın çok küçük bir oranını açıklamaktadır. Buna göre, görevlerin güçlük düzeylerinin birbirine yakın olduğu sonucuna varılmıştır. Hakem ana etkisi için kestirilen varyans bileşeninin toplam varyansın çok küçük bir oranını (%0.0-%3.7) açıkladığı görülmüştür. Bu da, hakemlerin tüm sporcuları puanlamasında, puanlamalar arasında tutarlılık olduğunu göstermektedir. Ancak sporcuhakem etkileşimine ilişkin varyans, toplam varyansı açıklamada ikinci sırada yer almaktadır. Bu da hakemlerin sporcuları, farklı katılık/cömertlikte puanladığı göstermektedir.

II. Buz pateni performansının ölçülmesinde hakem sayısının 12 olduğu yarışmalarda (2006-2008) G-katsayısının .98 ve 1.00 arasında, phi katsayısının ise .98 ve 1.00 arasında değiştiği görülmektedir. Karar çalışmasında hakem sayısının 9'a indirilmesi sonucunda G (.98 ve 1.00) ve phi (.98 ve .99) katsayılarında düşüş meydana gelmemiş, bağıl ve hata varyanslarında da çok az bir artış meydana gelmiştir. Hakem sayısının düşürülmesi genellenebilirlik ve güvenilirlik katsayılarında değişikliğe yol açmamaktadır.

III. 2009–2011 yıllarında elde edilen G ve phi katsayılarının da yüksek olduğu görülmektedir. Hakem sayısının 9 olduğu yarışmalarda G-katsayısının .98 ve .99 arasında, phi katsayısının ise yine .97 ve .99 arasında değiştiği görülmektedir. Karar çalışmaları açısından incelendiğinde ise 2009–2011 verilerinde hakem sayısının 12'ye çıkarılması artışa neden olmamaktadır. Bunun yanında, 2009–2011 yıllarında Dünya Buz Pateni Şampiyonalarında hakem sayısında azaltmaya gidilmesinin, genellenebilirlik ve güvenilirlik (Φ) katsayılarını olumsuz etkilemediği sonucuna varılmıştır. Ayrıca, elde edilen bulgulara göre, karar çalışmaları geçerli sonuçlar vermektedir.

IV. Çok değişkenlik kaynaklı Rasch ölçme modeline göre, 2006–2011 yıllarında yapılan Dünya Buz Pateni Şampiyonalarında, tekler (kadınlar ve erkekler) ve çiftler serbest programda, buz pateni performans ölçümlerinin model veri uyumuna FACET analizi ile bakılmış ve model veri uyumunun sağlandığı sonucuna varılmıştır.

V. Sporcuların performans ölçümlerine ilişkin yapılan çok değişkenlik kaynaklı Rasch ölçme modeli analizi sonucunda ayırma indeksi güvenilirliği ve ayırma oranı hesaplanmıştır. Bu değerler sporcu değişkenlik kaynağında oldukça yüksek olduğu bulunmuştur. Sporcu değişkenlik kaynağı söz konusu olduğunda; ayırma indeksinin güvenilirliği ve ayırma oranının yüksek olması arzu edilen bir durumdur. Sporcu değişkenlik kaynağı için ayırma indeksinin güvenilirliği Cronbach alfa ile benzemektedir (Myford ve Wolfe, 2003; Sudweeks, Reeve ve Bradshaw, 2005). Gerçek varyansın gözlenen varyansa oranını temsil etmektedir. 1'e yakın değerler yüksek güvenilirliği göstermektedir. Sporcu değişkenlik kaynağı söz konusu olduğunda bu değerlerin düşük olması, merkezi eğilim hatasını göstermektedir. Hakemlerin sporcuların performansını, söz konusu görevde ayırt edemediğini göstermektedir (Myford ve Wolfe, 2003; Sudweeks, Reeve ve Bradshaw, 2005). Ayırma indeksi güvenilirliği ve ayırma oranı (G) tüm yıllarda ve gruplarda oldukça yüksektir. Yani bir sporcunun yeteneği diğer sporcunun yeteneğinden ayrılmaktadır. Bu da merkezi eğilim hatasının sorun olmadığını göstermektedir.

VI. 2006–2008 yıllarındaki Dünya Buz Pateni yarışmalarında sporcuların performansının ölçülmesinde yer alan 12 hakemin, çok değişkenlik kaynaklı Rasch ölçme modeline göre hesaplanan ayırma indeksi güvenilirliğinin ve ayırma oranının

oldukça yüksek olduğu belirlenmiştir. Buna göre, hakemlerin cömertlik/katılık açısından birbirlerinden farklı oldukları sonucuna varılmıştır. Ancak, iç uyum ve dış uyum kareler ortalamasına ilişkin istatistiklerinin işaret ettiği şekliyle; bir bütün olarak hakemlerin tutarlı puanlama yapmış oldukları belirlenmiştir.

VII. 2009–2011 yıllarında yapılan Dünya Buz Pateni yarışmalarında sporcuların performansının ölçülmesinde yer alan 9 hakemin, çok değişkenlik kaynaklı Rasch ölçme modeline göre hesaplanan ayırma indeksi güvenilirliğinin ve ayırma oranının oldukça yüksek olduğu belirlenmiştir. Sadece 2010 erkekler serbest program (.02) ve 2011 çiftler serbest programda (.56) ayırma indeksi güvenilirlikleri arzu edildiği şekilde daha düşük bulunmuştur.

VIII. Buz pateni performansına ilişkin çok değişkenlik kaynaklı Rasch ölçme modeli analizlerine göre program bileşenlerinin (görevlerin) birbirinden farklı olduğu sonucuna varılmıştır. Buna göre, bu görevlerin her biri farklı yetenek düzeylerini gerektirmektedir.

IX. 2006–2011 yıllarında yapılan Dünya Buz Pateni Şampiyonalarında, tekler (kadınlar ve erkekler) ve çiftler serbest programda, sporcuların performansının ölçülmesine ilişkin olarak yapılan genellenebilirlik kuramı ve çok değişkenlik kaynaklı Rasch ölçme modeli analizleri sonucunda elde edilen istatistikler sporcu ve hakem değişkenlik kaynaklarına göre karşılaştırılmıştır. Sporcu boyutunda, iki kuramın karşılaştırılabilir değerleri olarak, genellenebilirlik kuramına göre genellenebilirlik katsayısının .98-1.00 ve çok değişkenlik kaynaklı Rasch ölçme modeline göre sporcu değişkenlik kaynağına ilişkin ayırma indeksi güvenilirliğinin .99-1.0 değerler arasında değiştiği bulunmuştur. Elde edilen bu değerlere göre ölçme sonuçlarının iç tutarlılık anlamında güvenilirliğinin oldukça yüksek olduğu sonucuna varılmıştır. Hakem boyutunda, tüm kuramların karşılaştırılabilir olan değerleri birlikte göz önüne alınmış ve şu sonuçlara varılmıştır: Genellenebilirlik kuramına göre, hakem değişkenlik kaynağına ilişkin kestirilen varyansın %0.0 - %3.4 arasında olduğu bulunmuştur. Elde edilen değerlerin, çok küçük olması ve tüm puanlara ilişkin genellenebilirlik katsayısının oldukça yüksek bir değer çıkması sonucunda, puanlar arasında hakemlerden kaynaklı bir değişimin gözlenmediği sonucuna varılmıştır. Çok değişkenlik kaynaklı Rasch ölçme modeline göre, hakem

değişkenlik kaynağına ilişkin ayırma indeksi güvenilirliği .02-.98 olarak hesaplanmıştır. Çoğunlukla yüksek olarak bulunan bu değer hakemlerin verdikleri puanlar arasında farklılık olduğunu göstermiştir. Ancak hakem değişkenlik kaynağına ilişkin uyum istatistiklerinin ortalamalarının 1'e yakın olması hakemlerin verdikleri puanlarda katılık/cömertlik açısından birbirleriyle tutarlı puanlama yaptıklarına işaret etmektedir.

X. Genellenebilirlik kuramında sporcuhakem etkileşimi, toplam varyansı açıklamada ikinci sırada yer almaktadır. Bu etkileşime ilişkin varyans, erkekler serbest programda %8.8-%12.4, kadınlar serbest programda %6.8-%14.4 ve çiftler serbest programda %2.8-%7.4 arasındadır. Sporcuhakem etkileşimi, belirli bir sporcuyu puanlamada, bazı hakemlerin katı, bazı hakemlerin ise cömert davrandığını göstermektedir. ÇDKRM'nde sporcuhakem etkileşimine ilişkin yanlılık analizi yapılmıştır. Bu analiz sonucuna göre, sporcuhakem etkileşimine ait yanlılık oranları; erkekler serbest programda %8.5-%11.5, kadınlar serbest programda %6.1-%13.7 ve çiftler serbest programda %2.8-%6.9'dur.

4.2 ÖNERİLER

Araştırma Sonuçlarına Göre Ortaya Çıkan Öneriler

I. Hakem sayısının azaltılması G ve Phi katsayılarında önemli bir değişikliğe neden olmamaktadır. Değerlendirmeler daha az hakemle yürütülebilir.

II. Genellenebilirlik kuramı ve ÇDKRM sundukları avantajlar bakımından, birbirini tamamlayıcı niteliktedir. Sporcu, hakem veya görevlere ilişkin değerlendirmeler bireysel olarak yapılmak istendiğinde ÇDKRM'nin kullanılması önerilirken, daha geniş kapsamlı değerlendirmelerin (bireysel ve grup düzeyinde, etkileşimleri dikkate alarak) yapılması için her iki tekniğin kullanılması yarar sağlayabilir.

İleride Yapılacak Çalışmalar İçin Öneriler:

I. Çalışma, hakemlerin kullanıldığı farklı spor dallarında (senkronize dalış, jimnastik) tekrarlanabilir.

II. Sporcu ve hakem sayıları değiştirilerek iki kuramı örneklem büyüklüğü açısından karşılaştıran araştırmalar yapılabilir.

III. Genellenebilirlik kuramındaki karar çalışmaları ile farklı koşullar (kısa/serbest program, tekler/çiftler) altında en uygun hakem sayılarını belirlemeye yönelik araştırmalar yapılabilir.

IV. ÇDKRM’de hakemler ve görevlerdeki farklılaşmanın nedenlerini (zevk, politik veya milliyetçi yönelimler, coğrafi bağlantı, yarışma sırası, yarışma düzeyi) belirlemeye yarayacak ileri düzey araştırmalar tasarlanabilir.

V. İki teknikte, hakem ve sporculara ilişkin elde edilen sonuçları, gözlem ve görüşme sonuçlarıyla karşılaştırmayı sağlayacak nicel ve nitel analiz tekniklerinin bir arada yer aldığı karma desenler kullanılarak yeni araştırmalar gerçekleştirilebilir.

KAYNAKLAR

- Aiken, L. R. (1996). *Rating scales and checklists: Evaluating behaviors, personality, and attitudes*. New York: John Wiley & Sons.
- Alharby, E. R. (2006). *A comparison between two scoring methods, holistic vs. analytic using two measurement models, the generalizability theory and the many facet rasch measurement within the context of performance assessment*. (Yayınlanmamış Doktora Tezi). The Pennsylvania State University, USA.
- American Educational Research Association, American Psychological Association ve National Council on Measurement in Education (1999). *Standarts for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillan.
- Anastasi, A. ve Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ.: Prentice Hall.
- Atılğan, H. (2004). *Genellenebilirlik kuramı ve çok değişkenlik kaynaklı rasch modelinin karşılaştırılmasına ilişkin bir araştırma* (Yayınlanmamış Doktora Tezi). Hacettepe Üniversitesi, Ankara.
- Atılğan, H. (2005). Genellenebilirlik kuramı ve puanlayıcılar arası güvenilirlik için örnek bir uygulama. *Eğitim Bilimleri ve Uygulama*, 4 (7).
- 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 in Education*, 31 (1), 63-76.
- Bachman, L.F., Lynch, B. K. ve Mason, M. (1995). Investigating variability in tasks and rater judgement in a performance test of foreign speaking. *Language Testing*, 12, 238-257.
- Baker, F. B. (1987). Methodology review: Item parameter estimation under the one-, two-, and three-parameter logistic models. *Applied Psychological Measurement*, 11, 111-142.

- Ball, D. W. (1973). A politicized social psychology of sport: Some assumptions and evidence from international figure skating competition. *International Review for the Sociology of Sport*, 8, 63-71.
- Basset, G. W. ve Persky, J. (1994). Rating skating. *Journal of American Statistical Association*, 89, 1075-1079.
- Baykul, Y. (2000). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması*. Ankara: ÖSYM.
- Bond, T. G. ve Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Bond, L. A., Braskamp, D. ve Roeber, E. (1996). *The status report of the assessment programs in the United States*. Oak Brook, IL: North Central Regional Educational Laboratory.
- Brennan, R. L. (1992). Generalizability theory. *Educational Measurement Issues and Practice*, 11 (4), 27-34.
- Brennan, R. L. (2001). *Statistics for social science and public policy: Generalizability theory*. New York: Springer-Verlag.
- Brown, J. D. (1991). Do English and ESL faculties rate writing samples differently? *TESOL Quarterly*, 25 (4), 587-603.
- Brown, G. T. L., Glasswell, K. ve Harland, D. (2004). Accuracy in the scoring of writing: Studies of reliability and validity using a New Zealand writing assessment system. *Assessing Writing*, 9, 105-121.
- Campbell, B. ve Galbraith, J. W. (1996). Non-parametric tests of the unbiasedness of figure-skating judgments. *Statistician*, 45 (4), 521-526.
- Cason, G. J. ve Cason, C. L. (1984). A deterministic theory of clinical performance rating. *Evaluation and the health professions*, 7(2), 221-247.
- Chang, W. ve Chan, C. (1995). Rasch analysis for outcome measures: Some methodological considerations. *Archives of Physical Medicine and Rehabilitation*, 76 (10), 934-939.

- Coffman, W. E. ve Kurfman, D. (1968). A comparison of two methods of reading essay examinations. *American Educational Research Journal*, 5 (1), 99-107.
- Conway, J. M. (1996). Analysis and design of multitrait-multirater performance appraisal studies. *Journal of Management*, 22 (1), 139-162.
- Cooper, W. H. (1981). Ubiquitous halo. *Psychological Bulletin*, 90 (2), 218-244.
- Crocker, L. ve Algina, J. (1986). *Introduction to classical and modern test theory*. Harcourt Brace Javanovich College Publishers, USA.
- Cronbach, L. J. (1990). *Essentials of psychological testing* (5th ed.). New York: Harper and Row.
- Cronbach, L. J., Gleser, G. C., Nanda, H. ve Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: Wiley.
- De Ayala, R.J. (1993). An introduction to polytomous item response theory models. *Measurement and Evaluation in Counselling and Development*, 3, 172-189.
- Deliceoğlu, G. (2009). *Futbol yetilerine ilişkin dereceleme ölçeğinin genellenebilirlik ve klasik test kuramına dayalı güvenilirliklerinin karşılaştırılması* (Yayınlanmamış Doktora Tezi). Ankara Üniversitesi, Ankara.
- Downing, S. (2003). Item response theory: Applications of modern test theory in medical education. *Medical Education*, 37 (8), 739-745.
- Downing, S. (2004). Reliability: On the reproductibility of assessment data. *Medical Education*, 38 (9), 1006-1012.
- Downing, S. M. ve Haladyna, T. M. (2004). Validity threats: Overcoming intereferece with proposed interpretations of assessment data. *Medical Education*, 38 (3), 327-333.
- Embretson, S. E. ve Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Engelhard, G. (1992). The measurement of writing ability with many-facet Rasch model. *Applied Measurement in Education*, 5 (3), 171-191.

- Engelhard, G. (1994). Examining rater errors in assessment of written composition with a many-faceted Rasch model. *Journal of Educational Measurement*, 31 (2), 93-112.
- Engelhard, G. (1996). Evaluating rater accuracy in performance assessment. *Journal of Educational Measurement*, 33 (1), 56-70.
- Engelhard, G. (2002). Monitoring raters in performance assessments. In G. Tindal and T. Haladyna (Eds.), *Large-scale assessment programs for all students: development, implementation, and analysis*, (pp. 261-287). Mahwah, NJ: Erlbaum.
- Fan, X. ve Chen, M. (2000). Published studies of interrater reliability often overestimate reliability: computing the correct coefficient. *Educational and Psychological Measurement*, 60 (4), 532-542.
- Feldt, L. S. ve Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 105-146). New York: American Council on Education and Macmillan.
- Findlay, L. ve Ste-Marie, D. (2004). A reputation bias in figure skating judging. *Journal of Sport and Exercise Psychology*, 26 (1), 154-163.
- Fisicaro, S. A., ve Lance, C. E. (1990). Implications of three causal models for the measurement of halo error. *Applied Psychological Measurement*, 14 (4), 419-429.
- Fitzpatrick, R., ve Morrisison, E. J. (1971). Performance and product evaluation. In R. L. Thorndike (Ed.), *Educational Measurement* (pp. 237-270). Washington, DC: American Council on Education.
- Fox, C. M. ve Jones, J. A. (1998). Uses of Rasch modeling in counseling psychology research. *Journal of Counseling Psychology*, 45 (1), 30-45.
- Furr, R.M. ve Bacharach, V.R. (2008). *Psychometrics: An Introduction*. Los Angeles: Sage Publication.

- Gao, X. ve Brennan, R. L. (2001). Variability of estimated variance components and related statistics in a performance assessment. *Applied Measurement in Education*, 14 (2), 191-203.
- Garicano, L., Palacios, I. ve Prendergast, C. (2001). Favoritism under social pressure. *NBER Working Paper*, no.8376.
- Garicano, L., Palacios, I. ve Prendergast, C. (2005). Favoritism under social pressure. *The Review of Economics and Statistics*, 87 (2), 208-216.
- Godbout, P. (1990). Observational strategies for the rating of motor skills: theoretical and practical implications. In M. Lirette, C. Pare, J. Dessureault & M. Pieron (Eds). *Physical education and coaching- present state and outlook for the future* Quebec: Quebec, Presses.
- Goodwin, L. D. (2001). Interrater agreement and reliability. *Measurement in Psychological Education and Exercises Science*, 5 (1), 13-14.
- Gould, D., Petlichkoff, L. ve Weinberg, R. S. (1984). Antecedents of temporal changes in and relationships between CSAI-2 subcomponents. *Journal of Sport Psychology*, 6 (3), 289-304.
- Gould, D., Petlichkoff, L., Simons, J. ve Vevera, M. (1987). Relationship between Competitive State Anxiety Inventory-2 subscale scores and pistol shooting performance. *Journal of Sport Psychology*, 9 (1), 33-42.
- Grehaigne, J.F. (1992). *The organization of play in soccer*. Paris: Actio.
- Grehaigne, J.F. ve Godbout, P. (1998). Formative assessment in team sports in a tactical approach context. *Journal of Physical Education, Recreation and Dance*, 69 (1), 46-51.
- Güler, N. (2008). *Klasik test kuramı, genellenebilirlik kuramı ve rasch modeli üzerine bir araştırma* (Yayınlanmamış Doktora Tezi). Hacettepe Üniversitesi, Ankara.
- Güler, N. ve Çetin, B. (2010). *Genellenebilirlik kuramı ve klasik test kuramı açısından gözetmen sayısı etkisinin incelenmesi*. Sözel Bildiri. II. Ulusal Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi, Mersin: Türkiye.

- Güler, N. ve Gelbal, S. (2010). Klasik test kuramı ve çok değişkenlik kaynaklı rasch modeli üzerine bir çalışma. *Eurasian Journal of Educational Research*, 38, 108-125.
- Haertel, E. H. (2006). Reliability. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 65-110). Westport, CT: American Council on Education and Praeger.
- Hambleton, R. (1989). Principles and selected applications of item response theory. In R. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 147-200). New York: Macmillan Publishing Company.
- Hambleton, R. ve Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Luwer-Nijhoff.
- Hambleton, R., Swaminathan, H. ve Rogers, H.J. (1991). *Fundamentals of item response theory*. Newbury Park: SAGE.
- Hellsten, L. M., Henderson, D. L. ve Klinger, D. A. (2000). *Using generalizability theory to examine the variation and irregularities in the scoring of figure skating performance*. Paper presented at annual meeting of the National Council on Measurement in Education. New Orleans, LA.
- Howell, D. C. (2002). *Statistical methods for psychology* (5th Edition). USA: Thomson Learning Academic Research Center.
- Hoyt, W. T. (2000). Rater bias in psychological research: When is it a problem and what can we do about it? *Psychological Methods*, 5(1), 64-86.
- Huang, J. ve Foote, C. J. (2011). Using generalizability theory to examine scoring reliability and variability of judging panels in skating competitions. *Journal of Quantitative Analysis in Sport*, 7 (3), Article 16.
- Iramaneerat, C. ve Yudkowsky, R. (2007). Rater errors in a clinical skills assessment of medical students. *Evaluation and the Health Profession*, 30(3), 266-283.
- ISU (2010). *Special regulations and technical rules*. <http://www.isu.org/vsite/vfile/page/fileurl/0,11040,4844-202745-219968-166758-0file,00.pdf> Erişim Tarihi: 03.09.2010

- Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp.17-64). Westport, CT: American Council on Education/Praeger.
- Keaveny, T. J. ve McGann, A. F. (1975). A comparison of behavioral expectation scales and graphic rating scales. *Journal of Applied Psychology*, 60 (6), 695-703.
- Kim, S.C. ve Wilson, M. (2009). a comparative analysis of the ratings in performance assessment using generalizability theory and many-facet rasch measurement. *Journal of Applied Measurement*, 10(4), 408-422.
- King, L. M., Hunter, J. E. ve Schmidt, F. L. (1980). Halo in a multidimensional forced-choice performance evaluation scale. *Journal of Applied Psychology*, 65(5), 507-516.
- Kondo-Brown, K. (2002). A FACETS analysis of rater bias in measuring Japanese second language writing performance. *Language Testing*, 19 (1), 3-31.
- Kozaki, Y. (2004). Using GENOVA and FACETS to set multiple standarts on performance assessment for certification in medical translation form Japanese to English. *Language Testing*, 21 (1), 1-27.
- Landy, F. J. ve Farr, J. L. (1980). Performance rating. *Psychological Bulletin*, 87, 72-107.
- Lee, J. (2004). Outlier aversion in evaluating performance: evidence from figure skating. *IZA Discussion Paper Series*, 1257.
- Lee, Y., Golub-Smith, M., Payton, C. ve Carey, J. (2001). *The score reliability of Test of Spoken English (TSE) from the generalizability theory perspective: Validating the current procedure*. Paper presented at the annual conference of American Educational Research Association (AERA), Seattle, WA.
- Lee, Y., Kantor, R. ve Mollaun, P. (2002). *Score dependability of the writing and speaking sections of new TOEFL*. Paper presented at the annual meeting of National Council on Measurement in education (NCME), New Orleans, LA.
- Lidbom, P. P. ve Priks, M. (2010). Behaviour under social pressure: Empty Italian stadiums and referee bias. *Economics Letters*, 108 (2), 212-214.

- Linacre, J.M. (1989). *Many-facet Rasch measurement* (Yayınlanmamış Doktora Tezi). University of Chicago, USA.
- Linacre, J.M. (1994). *Many-facet Rasch measurement*. Chicago: MESA.
- Linacre, J.M. (1995) Categorical misfit statistics. *Rasch Measurement Transactions*, 9, 3, 450-451. <http://www.rasch.org/rmt/rmt93j.htm> Erişim Tarihi: 26.07.2012.
- Linacre, J. M. (1997). Judging plans and facets. *MESA Research Note#3*. <http://www.Rasch.org/rn3.htm>. Erişim tarihi: 16.01.2012.
- Linacre, J.M. (2001). *A user's guide to Facets: Rasch measurement computer program*. Chicago: MESA.
- Linacre, J. M. (2002a). Judge ratings with forced agreement. *Rasch Measurement Transactions*, 16 (1), 857-858.
- Linacre, J. M. (2002b). What do infit and outfit mean-square and standardized mean? *Rasch Measurement Transactions*, 16 (2), 878.
- Linacre, J. M. (2007). *A User's Guide to Facets: Rasch Model Computer Programs*. Chicago, IL.
- Linacre, J. M. ve Wright, B. D. (2002). Construction of measures from many-facet data. *Journal of Applied Measurement*, 3 (4), 486-512.
- Linacre, J. M. ve Wright, B. D. (2004). Construction of measures from many-facet data. In E.V. Smith ve R.M. Smith (Eds.), *Introduction to Rasch Measurement* (pp.296-321). Maple Grove, MN: JAM Press.
- Linn, R. L. (1993). Educational assessment: Expanded expectations and challenges. *Educational Evaluation and Policy Analysis*, 15 (1), 1-16.
- Linn, R. L. ve Burton, E. (1994). Performance-based assessment: Implications of task specificity. *Educational Measurement: Issues and Practice*, 13 (1), 5-8, 15.
- Linn, R. L., Baker, E. L. ve Dunbar, S. B. (1991). Complex performance-based assessment: Expectations and validation criteria. *Educational Researcher*, 20 (8), 15-21.

- Looney, M. (1997). Objective measurement of figure skating performance. *Journal of Outcome Measurement*, 1 (2), 143-163.
- Looney, M. (2012). Judging anomalies at the 2010 Olympics in Men's Figure Skating. *Measurement in Physical Education and Exercise Science*, 16, 55-68.
- Lord, F. N. ve Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Lumley, T. ve McNamara, T. F. (1995). Rater characteristics and rater bias: implications for training. *Language Testing*, 12 (1), 54-71.
- Lunz, M. E. ve Schumacker, R. (1997). Scoring and analysis of performance examinations: a comparison of methods and interpretations. *Journal of Outcome Measurement*, 1 (3), 219-238.
- Lunz, M. E., Wright, B. D. ve Linacre, J. M. (1990). Measuring the impact of judge severity on examination scores. *Applied Measurement in Education*, 3(4), 331-345.
- Lynch, B. K. ve McNamara, T. F. (1998). Using G-theory and many-facet Rasch measurement in the development of performance assessments of the ESL speaking skills of immigrants. *Language Testing*, 15, 158-80.
- MacMillan, P. D. (1995). *A comparison of approaches: classical, generalizability and multifaceted Rasch* (Yayınlanmamış Doktora Tezi). Alberta University, Kanada.
- MacMillan, P. D. (2000). Classical, generalizability and multifaceted Rasch detection of interrater variability in large, sparse data sets. *The Journal of Experimental Education*, 68 (2), 167-190.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47 (2), 149-174.
- Matt, G. E. (2010). Generalizability theory. http://www.psychology.sdsu.edu/faculty/matt/Pubs/Gthml/Gtheory_GEMatt.htm
1. Erişim tarihi: 16/01/2012.
- McNamara, T. F. (1996). *Measuring second language performance*. New York: Longman.

- Murphy, K. R. (1982). Difficulties in the statistical control of halo. *Journal of Applied Psychology*, 67 (2), 161-164.
- Murphy, K. R. ve Balzer, W. K. (1989). Rater errors and rating accuracy. *Journal of Applied Psychology*, 74 (4), 619-624.
- Murphy, K. R. ve Cleveland, J. N. (1991). *Performance appraisal: an organizational perspective*. Massachusetts: Allyn and Bacon.
- Murphy, K.R ve Davidshofer, C. O. (2001). *Psychological testing: Principles and applications*. 5th Ed. New Jersey: Prentice Hall.
- Myford, C. M. ve Wolfe, E. W. (2002). When raters disagree, then what: examining a third-rating discrepancy resolution procedure and its utility for identifying unusual patterns of ratings. *Journal of Applied Measurement*, 3 (3), 300-324.
- Myford, C. M. ve Wolfe, E. W. (2003). Detecting and measuring rater effects using many-facet Rasch measurement: Part I. *Journal of Applied Measurement*, 4(4), 386-422.
- Myford, C.M., ve Wolfe, E.W. (2004) Detecting and measuring rater effects using many-facet Rasch measurement: Part II. *Journal of Applied Measurement*, 5(2), 189-227.
- Nadeau, L., Richard, J.F. ve Godbout, P. (2008). The validity and reliability of performance assessment procedure in ice hockey. *Physical Education and Sport Pedagogy*, 13 (1), 65-83.
- Nalbantođlu, F. (2009). *Performans ölçümlerinde genellenebilirlik kuramıyla farklı desenlerin karşılaştırılması* (Yayınlanmamış Yüksek Lisans Tezi). Hacettepe Üniversitesi, Ankara.
- Nande, P.J. ve Vali, S. A. (2010). *Fitness Evaluation tests for competitive sports*. India: Himalaya Publishing House.
- Nunally, J.C. ve Bernstein, I.H. (1994). *Psychometric Theory* (3rd ed.). New York: McGraw-Hill.
- O'Neill, T.R. (1999). Adjusting for rater severity over time. *Popular Measurement*, 47 (1), 46-47.

- Öztürk, M. E. (2011). *Voleybol becerileri ölçeğinin psikometrik özelliklerinin genellenebilirlik kuramı ve klasik test kuramına göre karşılaştırılması* (Yayınlanmamış Yüksek Lisans Tezi). Hacettepe Üniversitesi, Ankara.
- Park, T. (2004). An investigation of an ESL placement test of writing using many- facet Rasch measurement. <http://www.tc.columbia.edu/academic/tesol/webjournal/Taejoon2004.pdf> Erişim tarihi: 16/01/2012.
- Pedersen, G., Hagtvet, K. A. ve Katerud, S. (2007). Generalizability studied of the global assessment of functioning: Split version. *Comprehensive Psychiatry*, 48 (1), 88-94.
- Pinheiro, V. (1994). Diagnosing motor skills- a practical approach. *Journal of Physical Education, Recreation and Dance*, 65 (2), 49-54.
- Popham, W. J. (1990). *Modern educational measurement: A practitioner's perspective*. Englewood Cliffs, NJ: Prentice Hall.
- Ramineni, C. (2008). *Rater contrast effects in performance assessments using the medical licensure examination* (Yayınlanmamış Doktora Tezi). University of Delaware, USA.
- Raymond, M. R. ve Viswesavaran, C. (1993). Least squares models to correct for rater effects in performance assessment. *Journal of Educational Measurement*, 30 (3), 253-268.
- Rees, T., Hardy, L. ve Ingledew, D.K. (2000). Performance assessment in sport: Formulation, justification and confirmatory factor analysis of a measurement instrument for tennis performance. *Journal of Applied Sport Psychology*, 12 (2), 203-218.
- Resnick, D. P. ve Resnick, L. B. (1996). Performance assessment and the multiple functions of educational assessment. In M. B. Kane ve R. Mitchell (Eds.), *Implementing performance assessment: Promises, problems and challenges* (pp. 23-38), Mahwah, NJ: Erlbaum.
- Russell, M. K. ve Airasian, P. W. (2012). *Classroom assessment: Concepts and applications* (7. Ed.). New York: McGraw Hill.

- Saal, F., Downey, R. ve Lahey, M. A. (1980). Rating the ratings: assessing the quality of rating data. *Psychological Bulletin*, 88(2), 413-428.
- Saeki, N. ve Fan, X. (1999). *Scoring system reliability for international figure skating events: A generalizability analysis for 1998 winter Olympics*. Paper presented at the 46th Annual Meeting of the American College of Sports Medicine, Seattle, Washington.
- Scullen, S. E., Mount, M. K. ve Goff, M. (2000). Understanding the latent structure of job performance ratings. *Journal of Applied Psychology*, 85 (6), 956-970.
- Sebok, S. (2010). *"Pick me, pick me, i want to be a counsellor" assessment of med. counselling application selection process using Rasch analysis and generalizability theory* (Yayınlanmamış Yüksek Lisans Tezi). University of Northern British Columbia: USA.
- Seltzer, R. ve Glass, W. (1991). International politics and judging in Olympic skating events: 1968-1988. *Journal of Sport Behavior*, 14 (3), 189-200.
- Shavelson, R. J., Baxter, G. P. ve Gao, X. (1993). Sampling variability of performance assessments. *Journal of Educational Measurement*, 30 (3), 215-232.
- Shavelson, R. J. ve Webb, N. M. (1991). *Generalizability theory: A primer*. Newbury Park, CA: Sage.
- Shavelson, R. J., Webb, N. M. ve Rowley, G. L. (1989). Generalizability theory. *American Psychologist*. 44, 6, 922-932.
- Smith, Jr., E. V. ve Kulikowich, J. M. (2004). An application of generalizability theory and many-facet Rasch measurement using a complex problem-solving skills assessment. *Educational & Psychological Measurement*, 64, 617-639.
- Spears, J. F. (2008). *Design and utilization of performance assessment by vocational educators* (Yayınlanmamış Doktora Tezi). Illinois State University, USA.
- Stahl, J., Shumway, R., Bergstrom, B. ve Fisher, A. (1997). Online performance assessment using rating scales. *Journal of Outcome Measurement*, 1 (3), 173-191.

- Stemler, S. E. (2004). A comparison of consensus, consistency and measurement approaches to estimating interrater reliability. *Practical Assessment, Research and Evaluation*, 9 (4). <http://PAREonline.net/getvn.asp?=9&n=4> Erişim Tarihi: 16/01/2012.
- Sudweeks, R. R., Reeve, S. ve Bradshaw, W. S. (2005). A comparison of generalizability theory and many-facet Rasch measurement in an analysis of college sophomore writing. *Assessing Writing*, 9, 239-261.
- Swiss Society for Research in Education Working Group, (2010). *Edumetrics - Quality of measurement in education: EduG User Guide*, Neuchatel: Switzerland.
- Tritschler, K. (2000). *Barrow and McGee's practical measurement and assessment* (5th Ed.). Baltimore: Lippincott Williams and Wilkins.
- Turgut F. M. (1993). *Eğitimde ölçme ve değerlendirme metotları*. 9 baskı. Ankara: Saydam Matbaacılık.
- Turner, J. (2003). *Examining an art portfolio assessment using a many facet rasch measurement model* (Yayınlanmamış Doktora Tezi). Boston College, USA.
- Ure, A. C. (2011). *The effect of raters and rating conditions on the reliability of the missionary teaching assessment* (Yayınlanmamış Doktora Tezi).University of Brigham Young: USA.
- Viswesvaran, C., Ones, D. S. ve Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology*, 81 (5), 557- 574.
- Wallechinsky, D. (1991) *The complete book of the olympics*, (3rd ed.). New York: Little, Brown.
- Wanderer, J. J. (1987). Social factors in judges' ranking of competitors in figure skating. *Journal of Sport Behavior*, 10 (2), 93-102.
- Weekly, J.A. ve Gier, J.A. (1989). Ceilings in the reliability and validity of performance ratings: the case of expert raters. *The Academy of Management Journal*, 32 (1), 213-222.

- Weigle, S. C. (1998). Using Facets to model rater training effects. *Language Testing*, 15 (2), 263-287.
- Weinberg, R. S. (1990). Anxiety and motor performance: Where to go from here? *Anxiety Research*, 2, 227-242.
- Whissell, R., Lyons, S., Wilkinson, D. ve Whissell, C. (1993). National bias in judgements of Olympic-level of skating. *Perceptual and Motor Skills*, 77 (2), 355-358.
- Williamson, D. M., Mislevy, R. J. ve Bejar, I. I. (2006). *Automated scoring of complex tasks in computer based testing*. Mahwah, NJ: Erlbaum Associates.
- Wilson, M. ve Case, H. (2000). An examination of variation in rater severity over time: A study in rater drift. In M. Wilson, ve G. Engelhard (Eds.), *Objective measurement: Theory into practice* (Vol. 5, pp. 113-134). Stamford, CT: Ablex.
- Wilson, M. ve Wang, W. (1995). Complex composites: Issues that arise in combining different modes of assessment. *Applied Psychological Measurement*, 19 (1), 51-72.
- Wolfe, E. W. (2004). Identifying rater effects using latent trait models. *Psychology Science*, 46(1), 35-51.
- Wright, B. D. ve Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8 (3), 370.
- Wright, B. D. ve Linacre, J. M. (2001). Glossary of Rasch measurement terminology. *Rasch Measurement Transactions*, 15 (2), 824-825.
- Wu, S. ve Yang, M. (2004). Evaluation of the current decision rule in figure skating and possible improvements. *The American Statistician*, 58 (1), 46-55.
- Yamaguchi, K., Ness, C. ve Meacham, J. (1997). *Figure Skating for Dummies*. IDG Books Worldwide Inc.: Foster City, CA.
- Yang, R. (2010). *A many- facet rasch analysis of rater effecets on an oral english proficiency test* (Yayınlanmamış Doktora Tezi). Purdue University, USA.

- Yelboğa, A. (2007). *Klasik test kuramı ve genellenebilirlik kuramına göre güvenirliliğin bir iş performansı ölçęęi üzerinde incelenmesi* (Yayınlanmamış Doktora Tezi). Ankara Üniversitesi, Ankara.
- Yin, Y. ve Shavelson, R. J. (2008). Application of generalizability theory to concept map assessment research. *Applied Measurement in Education*, 21, 273-291.
- Zitzewitz, E. (2002). Nationalism in winter sport judging and its lesson for organizational decision making. Stanford GSB Research Paper No. 1796. <http://ssrn.com/abstract=319801> Erişim tarihi: 05.11.2011.
- Zitzewitz, E. (2006). Nationalism in winter sport judging and its lesson for organizational decision making. *Journal of Economics ve Management Strategy*, 15 (1), 67-99.

EK-1

2006- 2011 Dünya Buz Pateni Şampiyonası erkekler, kadınlar, çiftler serbest program sonuçlarına ilişkin G-çalışması sonuçları.

2006 Dünya Buz Pateni Şampiyonası Erkekler Serbest Program

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	Varyans	%
S	54876.964	23	2385.955	39.128	77.9
G	676.079	4	169.020	0.491	1.0
H	754.331	11	68.576	0.196	0.4
SG	1136.154	92	12.350	0.708	1.4
SH	7536.303	253	29.788	5.187	10.3
GH	838.954	44	19.067	0.634	1.3
SGH	3900.413	1012	3.854	3.854	7.7
Toplam	69719.197	1439			100%

2006 Dünya Buz Pateni Şampiyonası Kadınlar Serbest Program

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	Varyans	%
S	127203.483	23	5530.586	91.361	86.1
G	2139.190	4	534.798	1.787	1.7
H	2342.324	11	212.939	1.285	1.2
SG	574.910	92	6.249	0.330	0.3
SH	11379.492	253	44.978	8.538	8.0
GH	707.693	44	16.084	0.575	0.5
SGH	2317.407	1012	2.290	2.290	2.2
Toplam	146664.499	1439			100%

2006 Dünya Buz Pateni Şampiyonası Çiftler Serbest Program

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	Varyans	%
S	231404.487	19	12179.184	202.373	94.4
G	1545.728	4	386.432	1.540	0.7
H	1308.460	11	118.951	0.755	0.4
SG	489.805	76	6.445	0.307	0.1
SH	6917.773	209	33.099	6.067	2.8
GH	579.032	44	13.160	0.520	0.2
SGH	2310.235	836	2.763	2.763	1.3
Toplam	244555.520	1199			100%

EK-1 (DEVAM)**2007 Dünya Buz Pateni Şampiyonası Erkekler Serbest Program**

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	Varyans	%
S	96523.556	23	4196.676	69.139	82.6
G	3027.419	4	756.855	2.581	3.1
H	1669.656	11	151.787	0.852	1.0
SG	677.347	92	7.362	0.422	0.5
SH	10954.078	253	43.297	8.200	9.8
GH	373.747	44	8.494	0.258	0.3
SGH	2324.686	1012	2.297	2.297	2.7
Toplam	115550.489	1439			100%

2007 Dünya Buz Pateni Şampiyonası Kadınlar Serbest Program

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	Varyans	%
S	143301.556	23	6230.502	102.767	83.8
G	3594.315	4	898.579	3.035	2.5
H	1731.889	11	157.444	0.703	0.6
SG	923.618	92	10.039	0.486	0.4
SH	14835.311	253	58.638	10.885	8.9
GH	818.951	44	18.613	0.600	0.5
SGH	4261.515	1012	4.211	4.211	3.4
Toplam	169467.156	1439			100%

2007 Dünya Buz Pateni Şampiyonası Çiftler Serbest Program

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	Varyans	%
S	137057.408	17	8062.200	133.608	91.6
G	1236.189	4	309.047	1.382	0.9
H	457.047	11	41.550	-0.066	0.0
SG	377.411	68	5.550	0.267	0.2
SH	7945.136	187	42.487	8.029	5.5
GH	324.411	44	7.373	0.279	0.2
SGH	1751.989	748	2.342	2.342	1.6
Toplam	149149.592	1079			100%

EK-1 (DEVAM)**2008 Dünya Buz Pateni Şampiyonası Erkekler Serbest Program**

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	Varyans	%
S	132990.116	23	5782.179	95.254	81.7
G	3601.889	4	900.472	2.986	2.6
H	3608.974	11	328.089	2.063	1.8
SG	1421.311	92	15.449	0.955	0.8
SH	14032.809	253	55.466	10.296	8.8
GH	1278.911	44	29.066	1.045	0.9
SGH	4033.889	1012	3.986	3.986	3.4
Toplam	160967.899	1439			100%

2008 Dünya Buz Pateni Şampiyonası Kadınlar Serbest Program

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	Varyans	%
S	118821.948	22	5400.998	89.036	80.3
G	3362.010	4	840.503	2.909	2.6
H	6074.020	11	552.184	4.103	3.7
SG	862.523	88	9.801	0.539	0.5
SH	12678.730	242	52.391	9.812	8.8
GH	1375.781	44	31.268	1.215	1.1
SGH	3222.886	968	3.329	3.329	3.0
Toplam	146397.898	1379			100%

2008 Dünya Buz Pateni Şampiyonası Çiftler Serbest Program

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	Varyans	%
S	275609.349	19	14505.755	240.670	92.2
G	1665.245	4	416.311	1.656	0.6
H	4969.683	11	451.789	3.853	1.5
SG	792.555	76	10.428	0.631	0.2
SH	12116.001	209	57.971	11.022	4.2
GH	499.155	44	11.344	0.424	0.2
SGH	2390.245	836	2.859	2.859	1.1
Toplam	298042.233	1199			100%

EK-1 (DEVAM)**2009 Dünya Buz Pateni Şampiyonası Erkekler Serbest Program**

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	Varyans	%
S	83625.510	23	3635.892	79.156	78.8
G	2503.069	4	625.767	2.761	2.7
H	1701.035	8	212.629	1.075	1.1
SG	1053.643	92	11.453	0.904	0.9
SH	12093.765	184	65.727	12.481	12.4
GH	680.798	32	21.275	0.748	0.7
SGH	2443.291	736	3.320	3.320	3.3
Toplam	104101.110	1079			100%

2009 Dünya Buz Pateni Şampiyonası Kadınlar Serbest Program

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	Varyans	%
S	174202.441	23	7574.019	166.538	85.0
G	2528.015	4	632.004	2.687	1.4
H	7204.469	8	900.559	6.600	3.4
SG	1249.541	92	13.582	1.036	0.5
SH	12973.176	184	70.506	13.250	6.8
GH	1355.235	32	42.351	1.587	0.8
SGH	3134.009	736	4.258	4.258	2.2
Toplam	202646.885	1079			100%

2009 Dünya Buz Pateni Şampiyonası Çiftler Serbest Program

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	Varyans	%
S	148923.422	19	7838.075	172.870	90.8
G	899.144	4	224.786	1.115	0.6
H	1755.180	8	219.398	1.493	0.8
SG	622.322	76	8.188	0.526	0.3
SH	8240.998	152	54.217	10.152	5.3
GH	620.276	32	19.384	0.796	0.4
SGH	2102.658	608	3.458	3.458	1.8
Toplam	163164.000	899			100%

EK-1 (DEVAM)**2010 Dünya Buz Pateni Şampiyonası Erkekler Serbest Program**

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	Varyans	%
S	112061.907	23	4872.257	106.226	79.1
G	2924.939	4	731.235	3.301	2.5
H	200.963	8	25.120	-0.494	0.0
SG	1594.083	92	17.327	0.948	0.7
SH	15374.993	184	83.560	14.953	11.1
GH	309.611	32	9.675	0.037	0.0
SGH	6473.767	736	8.796	8.796	6.6
Toplam	138940.263	1079			100%

2010 Dünya Buz Pateni Şampiyonası Kadınlar Serbest Program

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	Varyans	%
S	117319.236	23	5100.836	110.690	77.5
G	3215.865	4	803.966	3.647	2.6
H	951.633	8	118.954	0.087	0.1
SG	1575.157	92	17.121	1.150	0.8
SH	20133.522	184	109.421	20.529	14.4
GH	188.219	32	5.882	-0.037	0.0
SGH	4985.959	736	6.774	6.774	4.7
Toplam	148369.592	1079			100%

2010 Dünya Buz Pateni Şampiyonası Çiftler Serbest Program

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	Varyans	%
S	120078.261	15	8005.217	176.514	91.2
G	1098.603	4	274.651	1.872	1.0
H	618.375	8	77.297	0.241	0.1
SG	379.753	60	6.329	0.319	0.2
SH	7106.114	120	59.218	11.152	5.8
GH	72.722	32	2.273	-0.074	0.0
SGH	1660.122	480	3.459	3.459	1.8
Toplam	131013.950	719			100%

EK-1 (DEVAM)**2011 Dünya Buz Pateni Şampiyonası Erkekler Serbest Program**

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	Varyans	%
S	91674.716	23	3985.857	86.727	76.6
G	3221.262	4	805.315	3.644	3.2
H	1111.227	8	138.903	0.576	0.5
SG	1835.961	92	19.956	1.284	1.1
SH	13173.773	184	71.597	12.640	11.2
GH	211.343	32	6.604	-0.075	0.0
SGH	6181.435	736	8.399	8.399	7.4
Toplam	117409.716	1079			100%

2011 Dünya Buz Pateni Şampiyonası Kadınlar Serbest Program

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	Varyans	%
S	121908.189	23	5300.356	116.577	88.7
G	1575.208	4	393.802	1.792	1.4
H	339.317	8	42.415	-0.081	0.0
SG	554.514	92	6.027	0.329	0.3
SH	9460.405	184	51.415	9.670	7.4
GH	120.000	32	3.750	0.029	0.0
SGH	2255.278	736	3.064	3.064	2.3
Toplam	136212.911	1079			100%

2011 Dünya Buz Pateni Şampiyonası Çiftler Serbest Program

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	Varyans	%
S	132768.410	15	8851.227	194.674	89.4
G	1250.714	4	312.678	2.103	1.0
H	367.028	8	45.878	-0.486	0.0
SG	605.153	60	10.086	0.648	0.3
SH	10207.328	120	85.061	16.161	7.4
GH	126.236	32	3.945	-0.019	0.0
SGH	2042.297	480	4.255	4.255	2.0
Toplam	147367.165	719			100%

EK-2

2006- 2011 Dünya Buz Pateni Şampiyonası erkekler, kadınlar, çiftler serbest program sonuçlarına ilişkin ÇDKRM sonuçları (sporcu yetenek, görev güçlük düzeyi ve hakem katılık düzeyi kestirimleri, iç uyum ve dış uyum değerleri).

2006 Dünya Buz Pateni Şampiyonası Erkekler Serbest Program/ Sporcu

Sporcu	Yetenek	Standart Hata	İç Uyum	Dış Uyum
1	3.01	.12	.87	.87
2	2.47	.12	1.42	1.41
4	2.38	.12	1.09	1.09
7	2.17	.12	.82	.81
3	1.46	.11	1.26	1.28
8	1.37	.11	.58	.60
9	1.37	.11	.58	.60
5	.84	.10	.58	.58
10	.42	.09	1.46	1.45
11	.08	.09	1.07	1.07
13	-.04	.10	1.07	1.05
15	-.33	.10	.46	.46
16	-.42	.10	.76	.74
18	-.81	.11	1.75	1.71
6	-.89	.11	1.44	1.39
14	-1.02	.11	.81	.82
12	-1.10	.11	.65	.64
20	-1.12	.11	.89	.90
22	-1.12	.11	.89	.90
23	-1.12	.11	1.35	1.34
21	-1.76	.11	.90	.90
19	-2.12	.11	.76	.78
17	-2.20	.10	1.70	1.76
24	-2.55	.10	.68	.72
Ortalama	-.04	.11	.99	.99
Standart Sapma	1.57	.01	.36	.36

2006 Dünya Buz Pateni Şampiyonası Erkekler Serbest Program/ Görev

Görev	Güçlük	Standart Hata	İç Uyum	Dış Uyum
Kayma Becerileri	.27	.05	.93	.92
Koreografi	.07	.05	.89	.89
Yorumlama	.00	.05	.98	.98
Performans	-.04	.05	.92	.93
Geçişler	-.29	.05	1.22	1.26
Ortalama	.00	.05	.99	.99
Standart Sapma	.18	.00	.12	.13

EK-2 (DEVAM)

2006 Dünya Buz Pateni Şampiyonası Erkekler Serbest Program/ Hakem

Hakem	Katılık	Standart Hata	İç Uyum	Dış Uyum
6	.35	.07	3.15	3.12
10	.24	.08	.84	.86
4	.21	.08	1.65	1.62
11	.07	.08	.60	.59
3	.07	.08	.86	.87
1	.00	.08	.43	.42
8	-.03	.08	.82	.84
12	-.06	.08	.47	.46
9	-.09	.07	.96	1.00
5	-.15	.07	.62	.63
2	-.26	.07	.45	.47
7	-.35	.07	1.04	1.04
Ortalama	.00	.07	.99	.99
Standart Sapma	.20	.00	.73	.72

2006 Dünya Buz Pateni Şampiyonası Kadınlar Serbest Program/ Sporcu

Sporcu	Yetenek	Standart Hata	İç Uyum	Dış Uyum
4	3.35	.10	.74	.74
2	3.21	.10	.94	.94
1	3.17	.10	.92	.93
3	2.66	.09	1.15	1.16
8	2.06	.09	1.43	1.45
6	1.95	.09	1.03	1.04
9	1.25	.09	.69	.68
5	1.24	.09	1.46	1.45
7	.97	.10	.76	.76
11	.95	.10	.55	.55
10	.47	.10	.66	.66
17	-.45	.09	.70	.70
12	-.81	.10	1.49	1.46
13	-.96	.10	1.01	1.02
15	-1.42	.10	1.11	1.07
18	-1.65	.10	1.61	1.57
14	-1.87	.11	.43	.43
20	-1.90	.11	1.03	1.03
21	-1.91	.11	.74	.75
24	-2.25	.11	1.09	1.08
16	-2.73	.11	1.27	1.27
22	-2.88	.11	.51	.52
19	-2.91	.11	1.10	1.11
23	-3.36	.11	1.17	1.17
Ortalama	-1.16	.10	.98	.98
Standart Sapma	2.13	.01	.32	.31

EK-2 (DEVAM)**2006 Dünya Buz Pateni Şampiyonası Kadınlar Serbest Program/ Görev**

Görev	Güçlük	Standart Hata	İç Uyum	Dış Uyum
Kayma Becerileri	.38	.05	.94	.93
Performans	.09	.05	.99	.98
Yorumlama	.04	.05	1.15	1.13
Koreografi	.02	.05	.96	.96
Geçişler	-.52	.05	.88	.89
Ortalama	.00	.05	.98	.98
Standart Sapma	.29	.00	.09	.08

2006 Dünya Buz Pateni Şampiyonası Kadınlar Serbest Program/ Hakem

Hakem	Katılık	Standart Hata	İç Uyum	Dış Uyum
4	.52	.07	1.13	1.16
6	.33	.07	1.85	1.88
8	.32	.07	1.56	1.53
12	.13	.07	.66	.66
11	.12	.07	.56	.56
5	.00	.07	.68	.67
1	-.08	.07	.87	.85
2	-.11	.07	1.42	1.44
7	-.14	.07	.32	.33
3	-.15	.07	.98	.99
10	-.27	.07	.58	.60
9	-.68	.07	1.17	1.09
Ortalama	.00	.07	.98	.98
Standart Sapma	.30	.00	.44	.44

EK-2 (DEVAM)**2006 Dünya Buz Pateni Şampiyonası Çiftler Serbest Program/ Sporcu**

Sporcu	Yetenek	Standart Hata	İç Uyum	Dış Uyum
1	6.78	.11	1.47	1.46
2	6.39	.12	1.00	.99
4	5.72	.11	.63	.64
5	5.68	.11	.74	.75
3	4.83	.11	.99	.98
7	4.73	.11	.94	.94
6	4.42	.11	1.31	1.32
8	3.15	.13	.89	.88
9	2.37	.12	.69	.65
10	.83	.10	.98	1.00
13	.78	.10	1.38	1.37
12	.30	.10	.45	.45
11	-1.01	.12	.86	.88
15	-2.77	.11	.87	.85
14	-3.48	.11	.86	.89
16	-3.55	.11	1.09	1.06
17	-3.84	.10	.92	.90
19	-4.00	.10	1.35	1.34
18	-4.39	.10	.68	.66
20	-6.35	.10	1.14	1.14
Ortalama	.83	.11	.96	.96
Standart Sapma	4.15	.01	.26	.26

2006 Dünya Buz Pateni Şampiyonası Çiftler Serbest Program/ Görev

Görev	Güçlük	Standart Hata	İç Uyum	Dış Uyum
Kayma Becerileri	.45	.05	.78	.77
Performans	.16	.05	.94	.93
Koreografi	.03	.05	.91	.92
Yorumlama	-.13	.05	1.11	1.09
Geçişler	-.50	.05	1.09	1.08
Ortalama	.00	.05	.97	.96
Standart Sapma	.30	.00	.12	.12

EK-2 (DEVAM)

2006 Dünya Buz Pateni Şampiyonası Çiftler Serbest Program/ Hakem

Hakem	Katılık	Standart Hata	İç Uyum	Dış Uyum
12	.63	.09	.77	.74
6	.34	.08	1.04	1.07
8	.16	.08	.60	.63
3	.12	.08	1.07	1.05
5	.08	.08	1.71	1.64
10	.06	.08	.85	.90
1	.04	.08	.86	.84
11	-.14	.08	.65	.63
9	-.22	.08	.57	.54
7	-.27	.08	.85	.89
2	-.29	.08	1.17	1.21
4	-.51	.08	1.45	1.34
Ortalama	.00	.08	.97	.96
Standart Sapma	.30	.00	.33	.31

2007 Dünya Buz Pateni Şampiyonası Erkekler Serbest Program/ Sporcu

Sporcu	Yetenek	Standart Hata	İç Uyum	Dış Uyum
2	4.18	.10	1.35	1.37
3	3.74	.10	.67	.67
1	3.69	.10	.80	.82
8	2.78	.11	.84	.84
4	2.65	.11	.76	.76
5	2.32	.11	.57	.58
10	2.27	.11	.81	.82
6	1.59	.11	1.12	1.13
14	.74	.11	.98	.97
16	.74	.11	.97	.97
7	.46	.11	1.09	1.08
9	.34	.11	1.07	1.06
11	.32	.11	.83	.83
12	.13	.11	.95	.96
15	-.04	.11	.92	.92
18	-1.04	.10	.77	.77
13	-1.11	.10	1.47	1.44
17	-1.36	.10	1.21	1.25
20	-1.36	.10	.85	.85
19	-1.91	.10	1.37	1.40
23	-2.22	.09	.49	.50
21	-2.26	.09	.93	.95
24	-3.04	.09	1.33	1.33
22	-3.07	.09	1.52	1.52
Ortalama	.36	.10	.99	.99
Standart Sapma	2.14	.01	.27	.27

EK-2 (DEVAM)**2007 Dünya Buz Pateni Şampiyonası Erkekler Serbest Program/ Görev**

Görev	Güçlük	Standart Hata	İç Uyum	Dış Uyum
Kayma Becerileri	.50	.05	.83	.81
Yorumlama	.13	.05	1.19	1.18
Performans	.03	.05	1.06	1.05
Koreografi	-.04	.05	1.04	1.02
Geçişler	-.63	.05	.90	.88
Ortalama	.00	.05	1.00	.99
Standart Sapma	.36	.00	.13	.13

2007 Dünya Buz Pateni Şampiyonası Erkekler Serbest Program/ Hakem

Hakem	Katılık	Standart Hata	İç Uyum	Dış Uyum
5	.47	.07	1.94	1.95
12	.42	.07	.64	.61
10	.35	.07	.90	.94
3	.10	.07	.99	.89
11	.01	.07	1.91	1.83
7	-.03	.07	.62	.60
2	-.10	.07	.57	.55
1	-.14	.07	1.10	1.16
9	-.16	.07	.78	.76
8	-.25	.07	.82	.83
6	-.26	.07	.95	.98
4	-.41	.07	.82	.80
Ortalama	.00	.07	1.00	.99
Standart Sapma	.27	.00	.44	.44

EK-2 (DEVAM)**2007 Dünya Buz Pateni Şampiyonası Kadınlar Serbest Program/ Sporcu**

Sporcu	Yetenek	Standart Hata	İç Uyum	Dış Uyum
1	3.47	.09	.90	.90
4	2.98	.09	.70	.69
2	2.63	.09	1.31	1.32
3	2.56	.09	1.36	1.37
9	2.10	.09	.85	.85
6	1.35	.09	.49	.50
8	1.05	.09	.64	.64
13	.96	.09	1.53	1.52
7	.62	.09	.84	.80
12	.56	.09	1.47	1.44
14	.35	.09	1.24	1.23
11	.28	.09	.81	.85
15	.28	.09	.81	.76
5	.14	.09	2.02	1.91
10	-.14	.08	1.14	1.11
16	-.41	.08	.38	.39
17	-.45	.08	.90	.89
20	-1.34	.08	1.68	1.76
23	-1.72	.08	.58	.58
24	-1.83	.08	.61	.61
21	-1.83	.08	.79	.79
18	-1.86	.08	.54	.54
19	-1.89	.08	.72	.72
22	-2.53	.08	1.27	1.23
Ortalama	.22	.09	.98	.98
Standart Sapma	1.68	.01	.41	.40

2007 Dünya Buz Pateni Şampiyonası Kadınlar Serbest Program/ Görev

Görev	Güçlük	Standart Hata	İç Uyum	Dış Uyum
Kayma Becerileri	.40	.04	.78	.78
Yorumlama	.06	.04	1.21	1.19
Performans	.03	.04	.94	.94
Koreografi	-.04	.04	.96	.96
Geçişler	-.45	.04	1.01	1.00
Ortalama	.00	.04	.98	.98
Standart Sapma	.27	.00	.14	.13

EK-2 (DEVAM)**2007 Dünya Buz Pateni Şampiyonası Kadınlar Serbest Program/ Hakem**

Hakem	Katılık	Standart Hata	İç Uyum	Dış Uyum
10	.30	.06	1.20	1.25
7	.22	.06	1.11	1.08
6	.15	.06	.77	.80
9	.12	.06	.54	.54
12	.08	.06	.35	.35
2	.06	.06	.38	.38
4	-.05	.06	.72	.70
5	-.05	.06	1.27	1.25
11	-.08	.06	1.02	1.02
8	-.11	.06	1.29	1.33
3	-.23	.06	.62	.57
1	-.39	.06	2.43	2.43
Ortalama	.00	.06	.97	.98
Standart Sapma	.19	.00	.54	.55

2007 Dünya Buz Pateni Şampiyonası Çiftler Serbest Program/ Sporcu

Sporcu	Yetenek	Standart Hata	İç Uyum	Dış Uyum
1	5.93	.12	.95	.95
2	2.63	.12	.79	.77
3	2.62	.12	.87	.85
4	.92	.10	1.79	1.79
10	.78	.10	.88	.90
5	.55	.10	.58	.60
6	.22	.10	.88	.88
8	.21	.10	.42	.42
7	.16	.10	1.10	1.12
9	-1.22	.09	.57	.58
11	-1.34	.09	1.17	1.16
13	-2.45	.09	.81	.80
16	-2.99	.10	.90	.90
15	-3.15	.10	.57	.57
12	-3.25	.10	1.03	1.02
14	-3.52	.10	1.04	1.03
18	-4.08	.10	1.39	1.39
17	-4.49	.10	.55	.55
Ortalama	-.69	.10	.90	.90
Standart Sapma	2.69	.01	.32	.32

EK-2 (DEVAM)**2007 Dünya Buz Pateni Şampiyonası Çiftler Serbest Program/ Görev**

Görev	Güçlük	Standart Hata	İç Uyum	Dış Uyum
Kayma Becerileri	.41	.05	.71	.69
Koreografi	.11	.05	1.12	1.09
Yorumlama	-.03	.05	.96	.95
Performans	-.12	.05	.90	.92
Geçişler	-.37	.05	.87	.87
Ortalama	.00	.05	.91	.90
Standart Sapma	.26	.00	.13	.13

2007 Dünya Buz Pateni Şampiyonası Çiftler Serbest Program/ Hakem

Hakem	Katılık	Standart Hata	İç Uyum	Dış Uyum
4	.24	.08	.94	.90
7	.13	.08	1.54	1.48
12	.11	.08	.52	.51
3	.09	.08	1.10	1.11
6	.07	.08	.43	.41
9	.07	.08	.59	.64
2	.05	.08	.46	.46
8	.00	.08	.84	.82
10	-.13	.08	.51	.54
5	-.14	.08	2.43	2.43
1	-.22	.08	.87	.89
11	-.28	.08	.67	.67
Ortalama	.00	.08	.91	.90
Standart Sapma	.15	.00	.55	.55

EK-2 (DEVAM)**2008 Dünya Buz Pateni Şampiyonası Erkekler Serbest Program/ Sporcu**

Sporcu	Yetenek	Standart Hata	İç Uyum	Dış Uyum
1	3.79	.11	1.00	.99
2	3.63	.11	1.37	1.31
7	3.01	.10	1.02	1.00
6	3.00	.10	1.00	.98
5	2.39	.09	.89	.87
20	1.56	.09	1.22	1.22
11	1.45	.09	1.62	1.64
9	1.13	.09	.83	.83
4	1.11	.09	.72	.73
3	1.04	.09	.57	.57
10	.98	.09	1.11	1.11
8	.58	.09	1.60	1.58
15	.48	.09	.78	.77
12	.15	.09	1.62	1.48
16	-.48	.08	.57	.57
13	-.62	.08	.67	.67
18	-.66	.08	1.63	1.59
14	-.71	.08	.45	.46
22	-.92	.08	.71	.71
19	-1.01	.08	.72	.72
17	-1.07	.08	.90	.93
21	-1.80	.07	.59	.61
23	-2.07	.07	1.38	1.38
24	-2.35	.07	1.02	1.03
Ortalama	.52	.09	1.00	.99
Standart Sapma	1.73	.01	.36	.35

2008 Dünya Buz Pateni Şampiyonası Erkekler Serbest Program/ Görev

Görev	Güçlük	Standart Hata	İç Uyum	Dış Uyum
Kayma Beceriler	.36	.04	.87	.84
Yorumlama	.09	.04	1.19	1.20
Koreografi	.03	.04	.92	.93
Performans	.01	.04	.96	.99
Geçişler	-.48	.04	.99	.99
Ortalama	.00	.04	.98	.99
Standart Sapma	.27	.00	.11	.12

EK-2 (DEVAM)

2008 Dünya Buz Pateni Şampiyonası Erkekler Serbest Program/ Hakem

Hakem	Katılık	Standart Hata	İç Uyum	Dış Uyum
6	.40	.06	1.94	1.87
8	.25	.06	.77	.74
1	.25	.06	.87	.89
4	.14	.06	.70	.74
12	.06	.06	.81	.85
9	.04	.06	.50	.51
11	.02	.06	.55	.54
3	-.02	.06	.40	.42
10	-.03	.06	.92	.84
7	-.05	.06	.90	.93
2	-.45	.06	1.51	1.47
5	-.62	.06	1.89	2.09
Ortalama	.00	.06	.98	.99
Standart Sapma	.27	.00	.50	.51

2008 Dünya Buz Pateni Şampiyonası Kadınlar Serbest Program/ Sporcu

Sporcu	Yetenek	Standart Hata	İç Uyum	Dış Uyum
2	3.47	.09	.77	.77
4	2.96	.09	.76	.80
3	2.88	.09	1.15	1.20
1	2.81	.09	1.31	1.28
6	2.10	.09	.94	.92
5	1.79	.09	.42	.42
17	.94	.09	1.87	1.93
12	.86	.09	1.63	1.62
8	.74	.09	1.08	1.09
15	.40	.09	.60	.61
7	.32	.09	.56	.58
13	.07	.09	.93	.99
11	-.21	.10	.62	.64
9	-.76	.10	.87	.87
14	-.94	.10	.73	.73
20	-1.09	.10	.68	.68
16	-1.44	.10	.89	.87
10	-1.59	.10	1.41	1.45
18	-1.60	.09	1.44	1.34
21	-1.87	.09	1.23	1.15
19	-1.93	.09	1.32	1.30
23	-2.68	.07	1.00	.88
22	-2.71	.07	.48	.50
Ortalama	.11	.09	.99	.98
Standart Sapma	1.85	.01	.37	.37

EK-2 (DEVAM)**2008 Dünya Buz Pateni Şampiyonası Kadınlar Serbest Program/ Görev**

Görev	Güçlük	Standart Hata	İç Uyum	Dış Uyum
Kayma Becerileri	.34	.04	.88	.88
Performans	.18	.04	1.07	1.07
Yorumlama	.06	.04	1.06	1.05
Koreografi	-.04	.04	.92	.92
Geçişler	-.54	.04	.99	1.00
Ortalama	.00	.04	.98	.98
Standart Sapma	.30	.00	.08	.07

2008 Dünya Buz Pateni Şampiyonası Kadınlar Serbest Program/ Hakem

Hakem	Katılık	Standart Hata	İç Uyum	Dış Uyum
12	.61	.07	.65	.65
9	.57	.07	.98	1.01
8	.49	.07	.72	.71
7	.25	.07	1.06	1.09
11	.18	.07	.95	.95
1	.06	.06	.56	.56
6	-.19	.06	1.21	1.17
4	-.28	.06	.61	.63
5	-.29	.06	.83	.85
10	-.29	.06	.74	.71
2	-.44	.06	1.20	1.24
3	-.66	.06	2.21	2.24
Ortalama	.00	.06	.98	.98
Standart Sapma	.40	.00	.43	.44

EK-2 (DEVAM)**2008 Dünya Buz Pateni Şampiyonası Çiftler Serbest Program/ Sporcu**

Sporcu	Yetenek	Standart Hata	İç Uyum	Dış Uyum
1	5.09	.10	1.11	1.13
5	4.09	.10	1.02	.99
3	3.84	.10	.72	.72
4	3.77	.10	1.18	1.21
2	3.35	.09	1.38	1.43
9	1.88	.09	1.59	1.57
6	1.65	.09	.63	.60
10	1.42	.08	1.42	1.27
7	1.28	.08	1.05	1.08
8	1.05	.08	1.06	1.05
11	.90	.08	.40	.40
12	.22	.07	.98	1.06
13	-.54	.08	1.05	1.11
14	-1.02	.09	.74	.77
15	-1.45	.10	.37	.38
17	-1.89	.10	1.12	1.12
16	-1.96	.10	.85	.84
19	-3.34	.09	.76	.74
18	-3.58	.09	1.11	1.05
20	-5.02	.08	.94	.94
Ortalama	.49	.09	.97	.97
Standart Sapma	2.74	.01	.31	.30

2008 Dünya Buz Pateni Şampiyonası Çiftler Serbest Program/ Görev

Görev	Güçlük	Standart Hata	İç Uyum	Dış Uyum
Kayma Becerileri	.31	.04	.80	.82
Performans	.07	.04	.94	.93
Koreografi	.07	.04	1.03	1.03
Yorumlama	-.09	.04	1.14	1.12
Geçişler	-.36	.04	.97	.96
Ortalama	.00	.04	.97	.97
Standart Sapma	.22	.00	.11	.10

EK-2 (DEVAM)

2008 Dünya Buz Pateni Şampiyonası Çiftler Serbest Program/ Hakem

Hakem	Katılık	Standart Hata	İç Uyum	Dış Uyum
5	.79	.07	.64	.61
7	.37	.07	1.61	1.71
8	.19	.07	1.09	1.08
1	.13	.07	.67	.69
3	.08	.07	1.49	1.47
2	.02	.07	.58	.63
10	-.01	.07	.41	.40
12	-.02	.07	.49	.50
11	-.08	.07	1.06	1.06
4	-.28	.07	1.16	1.18
9	-.30	.07	.89	.91
6	-.88	.07	1.57	1.43
Ortalama	.00	.07	.97	.97
Standart Sapma	.38	.00	.41	.40

2009 Dünya Buz Pateni Şampiyonası Erkekler Serbest Program/ Sporcu

Sporcu	Yetenek	Standart Hata	İç Uyum	Dış Uyum
1	2.27	.09	1.35	1.39
2	2.15	.09	1.59	1.61
3	1.92	.10	.47	.49
4	1.83	.10	.80	.80
5	1.59	.10	2.02	2.03
10	1.36	.10	1.35	1.35
7	1.00	.10	.88	.89
8	.78	.10	.74	.75
13	.00	.10	2.33	2.34
17	-.08	.10	.51	.51
12	-.19	.10	1.13	1.11
6	-.21	.10	.76	.77
15	-.44	.10	1.10	1.08
14	-.46	.10	.50	.49
9	-.69	.10	.42	.42
11	-.84	.09	.95	.95
16	-.86	.09	.65	.65
20	-1.03	.09	.88	.89
18	-1.40	.09	.79	.78
19	-1.58	.09	1.14	1.13
23	-1.99	.09	.74	.74
21	-2.29	.09	1.37	1.37
24	-2.51	.09	.64	.63
22	-2.62	.09	.63	.63
Ortalama	-.18	.10	.99	.99
Standart Sapma	1.48	.00	.48	.48

EK-2 (DEVAM)**2009 Dünya Buz Pateni Şampiyonası Erkekler Serbest Program/ Görev**

Görev	Güçlük	Standart Hata	İç Uyum	Dış Uyum
Kayma Becerileri	.30	.04	.94	.94
Yorumlama	.06	.04	1.13	1.14
Koreografi	.05	.04	.85	.84
Performans	.04	.04	.92	.94
Geçişler	-.46	.04	1.09	1.10
Ortalama	.00	.04	.98	.99
Standart Sapma	.25	.00	.11	.11

2009 Dünya Buz Pateni Şampiyonası Erkekler Serbest Program/ Hakem

Hakem	Katılık	Standart Hata	İç Uyum	Dış Uyum
9	.35	.06	.83	.79
6	.21	.06	.68	.66
5	.08	.06	.78	.76
7	.03	.06	.76	.76
8	.01	.06	.53	.52
3	-.05	.06	.47	.47
1	-.09	.06	1.04	1.08
2	-.13	.06	2.67	2.76
4	-.42	.06	1.09	1.13
Ortalama	.00	.06	.98	.99
Standart Sapma	.20	.00	.63	.66

EK-2 (DEVAM)**2009 Dünya Buz Pateni Şampiyonası Kadınlar Serbest Program/ Sporcu**

Sporcu	Yetenek	Standart Hata	İç Uyum	Dış Uyum
1	4.10	.11	.81	.85
2	2.99	.09	1.47	1.48
4	2.93	.09	1.02	1.02
3	2.83	.09	1.52	1.52
15	2.02	.10	1.59	1.55
7	2.00	.10	1.05	1.01
5	1.17	.10	.48	.49
10	1.06	.10	.57	.59
9	.90	.10	1.89	1.86
11	.89	.10	.57	.59
6	.76	.10	.54	.58
8	.68	.10	1.78	1.82
12	.60	.09	.66	.62
14	-.32	.08	1.24	1.21
13	-.64	.08	.78	.75
17	-.94	.08	.47	.48
19	-1.06	.08	.53	.54
21	-1.12	.08	.44	.45
20	-1.21	.08	.55	.56
16	-1.40	.08	.52	.52
18	-1.44	.08	.79	.78
22	-1.65	.09	2.43	2.40
24	-2.55	.09	.97	.96
23	-2.56	.09	1.03	1.04
Ortalama	.33	.09	.97	.96
Standart Sapma	1.81	.01	.53	.52

2009 Dünya Buz Pateni Şampiyonası Kadınlar Serbest Program/ Görev

Görev	Güçlük	Standart Hata	İç Uyum	Dış Uyum
Kayma Becerileri	.22	.04	.87	.86
Koreografi	.11	.04	.98	1.02
Yorumlama	.07	.04	1.12	1.16
Performans	.02	.04	.92	.94
Geçişler	-.42	.04	.96	.96
Ortalama	.00	.04	.97	.99
Standart Sapma	.22	.00	.08	.10

EK-2 (DEVAM)**2009 Dünya Buz Pateni Şampiyonası Kadınlar Serbest Program/ Hakem**

Hakem	Katılık	Standart Hata	İç Uyum	Dış Uyum
2	.43	.05	.87	.89
5	.33	.06	.57	.66
7	.33	.06	1.58	1.62
9	.30	.06	.47	.48
4	.27	.06	1.42	1.42
8	-.28	.06	2.09	2.05
1	-.32	.06	.33	.33
6	-.46	.05	.80	.79
3	-.60	.05	.64	.65
Ortalama	.00	.06	.97	.99
Standart Sapma	.38	.00	.55	.55

2009 Dünya Buz Pateni Şampiyonası Çiftler Serbest Program/ Sporcu

Sporcu	Yetenek	Standart Hata	İç Uyum	Dış Uyum
1	4.45	.12	.84	.76
3	2.14	.12	.92	.89
4	1.89	.11	1.49	1.36
2	1.88	.11	.89	.90
7	1.71	.11	.88	.91
6	.98	.09	1.10	1.05
5	.88	.09	.88	.84
8	.63	.09	1.23	1.32
9	.00	.09	.77	.73
12	-.12	.09	1.02	1.01
10	-.85	.10	.84	.84
11	-1.67	.10	.94	.93
14	-2.03	.10	1.28	1.32
13	-2.24	.11	1.40	1.42
15	-2.31	.11	.69	.67
16	-2.77	.11	1.21	1.20
17	-2.82	.11	1.16	1.16
18	-3.40	.11	.85	.85
19	-3.84	.11	.46	.46
20	-4.32	.11	.80	.80
Ortalama	-.59	.10	.98	.97
Standart Sapma	2.33	.01	.24	.25

EK-2 (DEVAM)**2009 Dünya Buz Pateni Şampiyonası Çiftler Serbest Program/ Görev**

Görev	Güçlük	Standart Hata	İç Uyum	Dış Uyum
Kayma Becerileri	.25	.05	.84	.84
Performans	.07	.05	1.02	1.01
Koreografi	.01	.05	.95	.92
Yorumlama	.00	.05	1.16	1.16
Geçişler	-.33	.05	.95	.92
Ortalama	.00	.05	.98	.97
Standart Sapma	.19	.00	.11	.11

2009 Dünya Buz Pateni Şampiyonası Çiftler Serbest Program/ Hakem

Hakem	Katılık	Standart Hata	İç Uyum	Dış Uyum
8	.47	.07	1.48	1.36
1	.38	.07	1.04	1.02
7	.10	.07	1.06	1.09
4	.00	.07	.52	.54
9	.00	.07	.30	.30
2	-.11	.07	1.17	1.13
6	-.15	.07	.89	.83
3	-.32	.07	1.31	1.36
5	-.38	.07	1.07	1.12
Ortalama	.00	.07	.98	.97
Standart Sapma	.27	.00	.35	.34

EK-2 (DEVAM)**2010 Dünya Buz Pateni Şampiyonası Erkekler Serbest Program/ Sporcu**

Sporcu	Yetenek	Standart Hata	İç Uyum	Dış Uyum
1	2.93	.11	1.15	1.18
2	2.18	.11	.84	.85
4	1.54	.10	1.45	1.37
6	1.01	.09	.36	.36
3	.99	.09	.92	.88
5	.85	.09	1.14	1.12
12	.63	.08	.48	.49
9	.26	.08	1.87	1.82
11	.23	.08	1.05	1.06
10	.19	.08	.86	.85
7	.19	.08	.88	.83
14	.00	.07	1.07	1.07
15	-.19	.07	.97	.96
13	-.20	.07	1.25	1.25
8	-.21	.07	1.29	1.29
19	-.36	.07	.60	.60
17	-.55	.07	1.30	1.28
16	-.95	.08	.76	.76
22	-1.05	.08	1.17	1.21
18	-1.16	.08	.99	.99
21	-1.31	.08	.57	.56
20	-1.44	.08	.94	.94
23	-1.82	.08	.65	.64
24	-2.05	.08	1.25	1.25
Ortalama	-.01	.08	.99	.98
Standart Sapma	1.19	.01	.33	.32

2010 Dünya Buz Pateni Şampiyonası Erkekler Serbest Program/ Görev

Görev	Güçlük	Standart Hata	İç Uyum	Dış Uyum
Kayma Becerileri	.19	.04	.83	.84
Yorumlama	.09	.04	1.09	1.05
Performans	.04	.04	.81	.80
Koreografi	.03	.04	.90	.88
Geçişler	-.35	.04	1.35	1.36
Ortalama	.00	.04	1.00	.98
Standart Sapma	.18	.00	.20	.21

EK-2 (DEVAM)**2010 Dünya Buz Pateni Şampiyonası Erkekler Serbest Program/ Hakem**

Hakem	Katılık	Standart Hata	İç Uyum	Dış Uyum
9	.10	.05	1.00	.88
7	.05	.05	.99	.96
6	.02	.05	1.27	1.29
4	.01	.05	.94	.94
8	.00	.05	.80	.78
5	-.03	.05	1.60	1.52
2	-.04	.05	1.01	1.05
1	-.05	.05	.69	.75
3	-.06	.05	.71	.69
Ortalama	.00	.05	1.00	.98
Standart Sapma	.05	.00	.27	.25

2010 Dünya Buz Pateni Şampiyonası Kadınlar Serbest Program/ Sporcu

Sporcu	Yetenek	Standart Hata	İç Uyum	Dış Uyum
1	2.16	.08	1.25	1.26
2	1.85	.08	1.08	1.09
5	1.66	.08	.51	.51
6	1.58	.08	1.54	1.54
3	1.26	.08	1.04	1.04
11	1.14	.08	.75	.75
4	1.09	.08	1.14	1.16
8	.67	.08	.83	.82
9	.64	.08	1.22	1.22
7	.46	.08	.99	.99
10	.08	.08	.91	.91
15	.06	.08	.76	.76
21	-.06	.07	1.30	1.29
14	-.14	.07	1.32	1.31
12	-.26	.07	.57	.56
20	-.34	.07	1.41	1.41
13	-.39	.07	.66	.66
16	-.67	.07	.83	.83
19	-.75	.07	.69	.69
17	-.76	.07	.56	.55
18	-1.08	.07	.83	.84
22	-1.24	.07	1.08	1.09
23	-1.41	.07	1.58	1.53
24	-1.73	.06	.88	.89
Ortalama	.16	.07	.99	.99
Standart Sapma	1.07	.00	.30	.30

EK-2 (DEVAM)

2010 Dünya Buz Pateni Şampiyonası Kadınlar Serbest Program/ Görev

Görev	Güçlük	Standart Hata	İç Uyum	Dış Uyum
Kayma Becerileri	.22	.03	.72	.74
Performans	.06	.03	1.01	1.01
Koreografi	.04	.03	.96	.96
Yorumlama	-.03	.03	1.16	1.16
Geçişler	-.30	.03	1.08	1.07
Ortalama	.00	.03	.99	.99
Standart Sapma	.17	.00	.15	.14

2010 Dünya Buz Pateni Şampiyonası Kadınlar Serbest Program/ Hakem

Hakem	Katılık	Standart Hata	İç Uyum	Dış Uyum
3	.14	.05	.99	.98
8	.13	.05	.80	.81
1	.06	.05	.97	.97
2	.03	.05	.91	.92
6	-.02	.05	1.25	1.21
4	-.02	.05	1.10	1.06
7	-.06	.05	.97	.98
5	-.12	.05	.97	.96
9	-.14	.05	.96	1.01
Ortalama	.00	.05	.99	.99
Standart Sapma	.09	.00	.12	.10

2010 Dünya Buz Pateni Şampiyonası Çiftler Serbest Program/ Sporcu

Sporcu	Yetenek	Standart Hata	İç Uyum	Dış Uyum
1	4.48	.13	1.01	1.00
2	4.25	.13	1.01	1.02
3	3.46	.11	.96	.95
4	3.28	.11	1.01	1.00
6	1.44	.11	.65	.65
5	1.35	.11	1.28	1.28
7	-.07	.11	.62	.61
8	-.40	.11	1.39	1.32
9	-1.04	.10	.75	.73
10	-1.05	.10	.83	.80
11	-1.84	.10	.84	.83
13	-2.12	.10	.89	.90
12	-2.64	.10	.88	.88
15	-3.16	.09	1.01	1.10
14	-3.20	.09	1.16	1.23
16	-3.24	.09	1.06	1.04
Ortalama	-.03	.11	.96	.96
Standart Sapma	2.65	.01	.20	.20

EK-2 (DEVAM)**2010 Dünya Buz Pateni Şampiyonası Çiftler Serbest Program/ Görev**

Görev	Güçlük	Standart Hata	İç Uyum	Dış Uyum
Kayma Becerileri	.31	.06	.81	.80
Performans	.10	.06	1.09	1.07
Koreografi	.04	.06	.93	.92
Yorumlama	-.06	.06	1.01	1.01
Geçişler	-.39	.05	.96	.99
Ortalama	.00	.06	.96	.96
Standart Sapma	.23	.00	.09	.09

2010 Dünya Buz Pateni Şampiyonası Çiftler Serbest Program/ Hakem

Hakem	Katılık	Standart Hata	İç Uyum	Dış Uyum
2	.38	.08	1.16	1.17
6	.18	.08	.73	.72
7	.05	.08	.65	.66
1	.01	.08	.56	.66
9	.01	.08	.95	.90
4	-.02	.08	1.15	1.11
3	-.16	.08	1.17	1.10
8	-.21	.07	1.05	.97
5	-.23	.07	1.18	1.33
Ortalama	.00	.08	.96	.96
Standart Sapma	.18	.00	.23	.23

EK-2 (DEVAM)**2011 Dünya Buz Pateni Şampiyonası Erkekler Serbest Program/ Sporcu**

Sporcu	Yetenek	Standart Hata	İç Uyum	Dış Uyum
1	2.50	.09	1.17	1.17
2	1.40	.08	.57	.57
6	1.35	.08	1.37	1.37
9	.89	.08	2.10	2.11
3	.87	.08	.62	.61
7	.73	.08	1.12	1.11
13	.65	.08	.77	.77
4	.56	.08	1.62	1.62
5	.47	.08	.56	.56
8	.34	.08	.80	.80
14	-.12	.08	.60	.60
12	-.18	.08	.81	.82
11	-.23	.08	1.04	1.04
17	-.30	.08	1.27	1.27
16	-.45	.08	.61	.61
10	-.56	.08	.63	.62
15	-.82	.08	.82	.81
21	-1.05	.08	1.07	1.09
18	-1.15	.08	1.13	1.13
19	-1.36	.09	.72	.71
20	-1.40	.09	1.57	1.58
24	-1.54	.09	.66	.66
22	-1.91	.09	1.09	1.09
23	-2.05	.09	.98	.99
Ortalama	-.14	.08	.99	.99
Standart Sapma	1.13	.00	.38	.39

2011 Dünya Buz Pateni Şampiyonası Erkekler Serbest Program/ Görev

Görev	Güçlük	Standart Hata	İç Uyum	Dış Uyum
Kayma Becerileri	.19	.04	.81	.81
Yorumlama	.12	.04	1.18	1.18
Performans	.07	.04	1.09	1.10
Koreografi	.04	.04	.88	.88
Geçişler	-.42	.04	.96	.97
Ortalama	.00	.04	.98	.99
Standart Sapma	.22	.00	.14	.13

EK-2 (DEVAM)

2011 Dünya Buz Pateni Şampiyonası Erkekler Serbest Program/ Hakem

Hakem	Katılık	Standart Hata	İç Uyum	Dış Uyum
9	.24	.05	.82	.81
3	.17	.05	.83	.84
7	.07	.05	.98	.99
5	-.02	.05	.88	.90
6	-.04	.05	1.05	1.07
4	-.08	.05	.97	.97
2	-.09	.05	1.27	1.25
8	-.10	.05	.90	.91
1	-.16	.05	1.14	1.15
Ortalama	.00	.05	.98	.99
Standart Sapma	.13	.00	.14	.14

2011 Dünya Buz Pateni Şampiyonası Kadınlar Serbest Program/ Sporcu

Sporcu	Yetenek	Standart Hata	İç Uyum	Dış Uyum
2	4.50	.11	.81	.80
3	3.80	.11	.94	.94
1	3.71	.11	.81	.81
4	3.05	.11	1.55	1.57
5	2.80	.11	1.41	1.42
9	2.63	.11	1.16	1.15
6	2.45	.11	.91	.92
8	1.86	.12	.74	.74
7	1.42	.12	.71	.71
12	.18	.12	1.26	1.19
14	-.22	.11	1.77	1.77
10	-.36	.10	.63	.64
11	-.55	.10	.56	.55
18	-.92	.10	1.40	1.40
15	-1.11	.10	.86	.86
16	-1.24	.10	1.01	1.02
13	-1.29	.10	.85	.85
19	-1.91	.11	.68	.71
17	-2.21	.12	.75	.75
20	-2.26	.12	.80	.80
21	-2.33	.12	1.10	1.08
24	-2.47	.12	.93	.93
22	-2.80	.12	.82	.80
23	-3.84	.10	1.14	1.14
Ortalama	-.12	.11	.98	.98
Standart Sapma	2.39	.01	.30	.30

EK-2 (DEVAM)

2011 Dünya Buz Pateni Şampiyonası Kadınlar Serbest Program/ Görev

Görev	Güçlük	Standart Hata	İç Uyum	Dış Uyum
Kayma Becerileri	.24	.05	.86	.86
Yorumlama	.15	.05	1.08	1.05
Performans	.07	.05	.97	.95
Koreografi	.04	.05	.90	.90
Geçişler	-.50	.05	1.14	1.14
Ortalama	.00	.05	.99	.98
Standart Sapma	.26	.00	.11	.10

2011 Dünya Buz Pateni Şampiyonası Kadınlar Serbest Program/ Hakem

Hakem	Katılık	Standart Hata	İç Uyum	Dış Uyum
2	.22	.07	1.09	1.09
6	.12	.07	1.03	1.03
8	.09	.07	1.07	1.07
7	.00	.07	.75	.75
1	-.03	.07	.88	.83
9	-.05	.07	.98	.97
3	-.07	.07	1.40	1.42
4	-.08	.07	.69	.70
5	-.21	.07	1.01	.98
Ortalama	.00	.07	.99	.98
Standart Sapma	.12	.00	.20	.20

2011 Dünya Buz Pateni Şampiyonası Çiftler Serbest Program/ Sporcu

Sporcu	Yetenek	Standart Hata	İç Uyum	Dış Uyum
1	3.95	.11	1.09	1.10
2	3.46	.11	1.23	1.23
3	2.93	.11	.65	.65
4	1.93	.10	.65	.65
5	1.10	.09	1.55	1.52
6	.52	.09	.89	.89
10	-.31	.09	.71	.71
7	-.31	.09	1.31	1.31
8	-.55	.09	.79	.79
9	-.91	.09	1.08	1.08
11	-1.22	.09	.83	.82
12	-1.27	.08	1.05	1.05
13	-1.94	.08	1.41	1.40
14	-2.14	.08	1.04	1.01
15	-2.17	.08	.47	.47
16	-2.74	.07	.99	.99
Ortalama	.02	.09	.98	.98
Standart Sapma	2.04	.01	.29	.28

EK-2 (DEVAM)**2011 Dünya Buz Pateni Şampiyonası Çiftler Serbest Program/ Görev**

Görev	Güçlük	Standart Hata	İç Uyum	Dış Uyum
Kayma Becerileri	.19	.05	.83	.83
Koreografi	.08	.05	.96	.96
Yorumlama	.07	.05	.97	.97
Performans	-.01	.05	1.03	1.04
Geçişler	-.33	.05	1.12	1.11
Ortalama	.00	.05	.98	.98
Standart Sapma	.18	.00	.09	.09

2011 Dünya Buz Pateni Şampiyonası Çiftler Serbest Program/ Hakem

Hakem	Katılık	Standart Hata	İç Uyum	Dış Uyum
2	.16	.07	.81	.89
6	.08	.07	1.06	.93
5	.05	.07	.53	.54
1	.04	.07	1.05	1.07
8	.03	.07	.46	.47
7	-.02	.07	1.20	1.30
9	-.03	.07	1.14	1.07
4	-.15	.06	1.65	1.51
3	-.16	.06	.94	1.04
Ortalama	.00	.07	.98	.98
Standart Sapma	.10	.00	.34	.31