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

FİZYOTERAPİ VE REHABİLİTASYON ÖĞRENCİLERİNDE EKLEM HAREKET ÖLÇÜM BECERİLERİNİN PRATİK DEĞERLENDİRİLMESİNİN GENELLENEBİLİRLİK KURAMINA GÖRE İNCELENMESİ

YÜKSEK LİSANS TEZİ

HAZIRLAYAN
YAVUZ YAKUT

GAZİANTEP – 2021

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

FİZYOTERAPİ VE REHABİLİTASYON ÖĞRENCİLERİNDE EKLEM HAREKET ÖLÇÜM BECERİLERİNİN PRATİK DEĞERLENDİRİLMESİNİN GENELLENEBİLİRLİK KURAMINA GÖRE İNCELENMESİ

YÜKSEK LİSANS TEZİ

HAZIRLAYAN
YAVUZ YAKUT

TEZ DANIŞMANI
PROF. DR. ŞENER BÜYÜKÖZTÜRK

GAZİANTEP – 2021

SOSYAL BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜNE
YÜKSEK LİSANS KABUL VE ONAY FORMU

Eğitimde Ölçme ve Değerlendirme Anabilim Dalı **Eğitimde Ölçme ve Değerlendirme** Tezli Yüksek Lisans Programı öğrencisi **Yavuz YAKUT** tarafından hazırlanan **"Fizyoterapi ve Rehabilitasyon Öğrencilerinde Eklem Hareket Ölçüm Becerilerinin Pratik Değerlendirilmesinin Genellenebilirlik Kuramına Göre İncelenmesi"** başlıklı tez, **21/ 06 / 2021** tarihinde yapılan savunma sınavı sonucu **başarılı** bulunarak jürimiz tarafından **Yüksek Lisans Tezi** olarak kabul edilmiştir.

Görevi	**Unvanı, Adı ve Soyadı Kurumu/Üniversitesi**	**İmzası:**
Tez Danışmanı	Prof. Dr. Şener BÜYÜKÖZTÜRK Hasan Kalyoncu Üniversitesi	
Jüri Üyesi	Doç. Dr. Ufuk AKBAŞ Hasan Kalyoncu Üniversitesi	
Jüri Üyesi	Dr. Öğr. Üyesi Ersoy KARABAY Kırşehir Ahi Evran Üniversitesi	

Bu tez Enstitü Yönetim Kurulunca belirlenen yukarıdaki jüri üyeleri tarafından uygun görülmüş ve Enstitü Yönetim Kurulu kararı ile onaylanmıştır.

.....
Enstitü Müdürü

TEZ ETİK VE BİLDİRİM SAYFASI

"Yüksek lisans tezi olarak sunduğum **"FİZYOTERAPİ VE REHABİLİTASYON ÖĞRENCİLERİNDE EKLEM HAREKET ÖLÇÜM BECERİLERİNİN PRATİK DEĞERLENDİRİLMESİNİN GENELLENEBİLİRLİK KURAMINA GÖRE İNCELENMESİ"** başlıklı çalışmanın tarafımca, bilimsel ahlak ve geleneklere aykırı düşecek bir yardıma başvurmaksızın yazıldığını ve yararlandığım eserlerin kaynakçada gösterilenlerden oluştuğunu ve bunlara atıf yapılarak yararlanmış olduğumu belirtir ve onurumla doğrularım." 28.06.2021

Yavuz YAKUT

ÖNSÖZ

Tez çalışmam boyunca değerli katkı ve yorumlarıyla beni yönlendiren danışmanım Prof. Dr. Şener Büyüköztürk'e teşekkürlerimi sunarım.

Bilim uzmanlığı eğitimim süresince desteğini esirgemeyen hocam, Doç. Dr. Ufuk Akbaş'a,

Değerli katkı ve desteklerinden dolayı Arş. Gör. Merve Yıldırım Seheryeli'ne,

Tezin gerçekleşmesindeki katkı ve emeklerinden dolayı, değerli arkadaşım Öğr. Gör. Dilek Yamak'a, sevgili meslektaşım Dr. Öğ. Üyesi Duygu Türker'e,

Tezin gerçekleştirilmesindeki katkılarından dolayı Prof. Dr. Kezban Bayramlar'a, Prof. Dr. Necmiye Ün Yılıdırım'a, çok sevgili arkadaşım Prof. Dr. Nur Tunalı'ya, Doç. Dr. Naciye Vardar Yağlı'ya, Yrd. Doç. Dr. Sevim Öksüz'e, Dr. Öğr. Üyesi Deniz Erdan Kocamaz'a,

Arş. Gör. Zeynep İrem Bulut'a, Arş. Gör. Ekin Oğuz Sarı'ya, Arş. Gör. Merve Karatel'e,

Özveri ve ciddiyetle çalışmaya destek veren çok değerli öğrencilere ve ismini sayamadığım dostlarıma, emeği geçen herkese,

Teşekkür ederim.

ÖZET

Bu çalışmada, fizyoterapi ve rehabilitasyon öğrencilerinde eklem hareket ölçüm becerilerinin pratik değerlendirilmesinin genellenebilirlik kuramına göre incelenmesi amaçlanmıştır. Çalışmada fizyoterapi ve rehabilitasyon öğrencilerinde eklem hareket performansı 6 farklı öğretim üyesi tarafından değerlendirilmiştir. Her puanlayıcı farklı zamanlarda dereceleme ölçeği ve genel izlenim ile performans değerlendirmesi yapmıştır. Dereceleme ölçeği ve genel izlenim performans değerlendirmelerinin test-tekrar test güvenirliği ve iki farklı puanlamanın puanlayıcılar arası ilişkisi incelenmiştir. Dereceleme ölçeği puanlama sonuçları Klasik Test Kuramı ve Genellenebilirlik Kuramına göre analiz edilmiş ve karşılaştırılmıştır. Genellenebilirlik çalışması birey, puanlayıcı ve madde çapraz desenine göre gerçekleştirilmiş (b x p x m) ve Karar çalışması yapılmıştır. Çalışma sonucuna göre, her iki puanlama türünün test-tekrar test güvenirliği mükemmel bulunmuştur. Altı farklı puanlayıcı arasında dereceleme ölçeği ve genel izlenim puanlamalarında sınıfiçi korelasyon katsayısı yüksek çıkmasına rağmen, puanlayıcılar arası farkların olduğu belirlenmiştir. Dereceleme ölçeği performans değerlendirme sonuçlarında, Klasik Test Kuramı ve Genellenebilirlik Kuramı katsayıları yüksek çıkmıştır. Genellenebilirlik Kuramı analizinde maddelerin puanlamayı etkilemediği, birey ve puanlayıcının ve birey-puanlayıcı bileşeninin daha etkili olduğu görülmüştür. Artık etkinin en yüksek yüzdeye sahip olması, puanlamayı etkileyen yorgunluk, farklı performansların karşılaştırılması gibi çalışmaların yapılmasının yararlı olacağını göstermiştir.

Anahtar kelimeler: Klasik Test Kuramı, Genellenebilirlik Kuramı, Performans değerlendirme, Güvenirlik, Dereceleme ölçeği, Genel izlenim.

ABSTRACT

In this study, it is aimed to examine the practical evaluation of joint motion measurement skills in physiotherapy and rehabilitation students according to the generalizability theory. Performance of joint motion measurements in physiotherapy and rehabilitation students were evaluated by six different faculty members. Each scorer made rating scale and general impression performance evaluation at different times. Test-retest reliability of rating scale and general impression rating scale and general impression rating scale and general impression performance evaluations and the relationship between raters of two different scoring was examined. Rating scale results were analyzed and compared according to Classical Test Theory (CTT) and Generalizability Theory (GT). The generalizability study was carried out according to the student, rater, and item cross design (s x r x i) and the Decision study was conducted. According to the results of the study, the test-retest reliability of both scoring types was found to be excellent. Although the intra-class correlation coefficient was high in rating scale and general impression scoring among six different raters, it was determined that there were differences between raters. In rating scale performance evaluation results, the CTT and the GT coefficients were found to be high. In the GT analysis, it was observed that the items did not affect the scoring, and the individual, rater, and the individual-rater component were more effective. It has shown that it would be beneficial to conduct studies such as having the highest percentage of residual effect, fatigue affecting the scoring, and comparison of different performances.

Keywords: Classical test theory, Generalizability theory, Performance assessment, Reliability, Rating scale, General impression.

İÇİNDEKİLER

Sayfa No.

TABLOLAR LİSTESİ

Sayfa No.

ŞEKİLLER LİSTESİ

Sayfa No.

BİRİNCİ BÖLÜM

GİRİŞ

Giriş bölümünde, fizyoterapi ve rehabilitasyon eğitimi alanında önemli bir yer tutan pratik performansın değerlendirilmesi temelinde problemin durumu, çalışmanın hedeflerine değinilmiştir. Sağlık alanında performansın değerlendirilmesinde kullanılan yöntemlerden ve özelliklerinden bahsedilmişti. Temel araştırma hedefi ve belirlenen alt hedeflere yer verilmiştir.

1.1. Problem Durumu

Sağlık alanında mesleki bilgi ve becerilerin ölçülmesi ve değerlendirilmesinde, sürekli yenilenme ve gelişme zorunluluğu bulunmaktadır. Bilginin sürekli gelişimi, eğitim veren kurumların ve eğitim alan öğrenci sayılarındaki aşırı artışlar, beraberinde standardizasyon sorunlarını getirmektedir. Öğretilen bilgi ve öğrenilen performansın değerlendirilmesinde klasik yöntemlerin kullanılması, çoğunlukla öğrenme eksikliklerinin belirlenememesine, öğretim aşamasındaki eksiklik ve geliştirilmesi gereken yönlerin değerlendirilememesine neden olmaktadır. Bir performans içerisindeki farklı görevlerin birbirinden bağımsız olarak analiz edilebilmesi, performansı oluşturan temel öğrenim öğelerini ayrı ayrı değerlendirebilme fırsatı vermektedir. Bu sayede, sağlık alanında eğitim alan öğrencilerin gerçek hasta üzerinde yapacakları uygulamalara daha iyi ve donanımlı olarak hazırlanabilmeleri sağlanmaktadır.

Farklı görevleri içeren bir pratik performans değerlendirilmesinde temel sorunlar görev tanımları, puanlama güvenirliği ve tutarlılığıdır. Bu durumun belirlenmesinde uzun yıllardır klasik test kuramına göre değerlendirme ve analizler yapılmaktadır. Son yıllarda kullanımı giderek artan Genellenebilirlik Kuramının daha kapsamlı değerlendirme ve incelemeleri mümkün kılmaktadır. Ancak, fizyoterapi ve rehabilitasyon alanında genellenebilirlik kuramına göre performans değerlendirme örnekleri bulunmamaktadır.

1.1.1. Eğitimde değerlendirme

Eğitimde değerlendirme genel bir ifadedir. Değerlendirmenin, eğitim sürecindeki etkisinin yanında, öğrenci ve değerlendiricilerin eğitim ve değerlendirme süreçlerindeki rolüne bağlı olarak değişiklik göstermektedir (Birenbaum, 1994). Değerlendirme kavramındaki en önemli değişiklik, 'bir öğrenme aracı olarak değerlendirme' kavramıyla açıklanabilir (Dochy ve McDowell 1997). Geçmişte öğrenci değerlendirmede not verme işlemi, öğrencinin amaçlanan hedefe ne ölçüde ulaştığını belirlemek amacıyla kullanılan bir araç olarak görülmüştür. Günümüzde, değerlendirmenin olası kazanımlarının çok daha geniş olduğu ve öğrenme sürecinin tüm aşamalarını etkilediği kabul edilmektedir (van de Watering ve diğerleri, 2008).

Öğrencinin öğrendiğinin analiz edilmesi her zaman için eğitimin önemli bir parçası olmuştur. 20. yüzyılın sonlarından itibaren, eğitimin her aşamasında öğrencilerin nasıl değerlendirileceklerine yönelik yenilikçi ve sorumluluğun arttığı bir yaklaşıma yönelim olmuştur (Vendrely, 2002). Eğiticiler için, öğrencinin her alanda uygun değerlendirilmesi önemli bir konudur. Hedeflenen öğrenim çıktılarının elde edilmesinde öğretim yollarına uygun şekilde öğrencinin değerlendirilmesi esastır. Değerlendirme aşağıda belirtilen konulardan bir ya da birkaçını içermelidir: Öğrencinin bireysel öğrendikleri, öğrencinin gelişimi veya eksik yönleri, belirli bir alanda öğrencinin yeterlilik veya becerisinin geliştirilmesi, diğer gruplara göre öğrencinin kazanımlarının karşılaştırılması. Değerlendirme fizyoterapi ve rehabilitasyon öğretim programında hayati derecede öneme sahiptir. Öğrenciyi motive edebilir, hatalarını düzeltebilir, edinilen bilgiyi ölçebilir, eğitimin etkilerini değerlendirebilir, uygun olmayan öğrenci davranışlarını belirleyebilir, iyileştirmeye yönelik belirli planların geliştirilmesine yardım eder (Cross, 1983; Cross ve Hicks, 1997).

Dünya Fizyoterapi Konfederasyonu (The World Confederation of Physical Therapy (WCPT)), tüm fizyoterapi eğitim programlarında içeriğin yaklaşık üçte birinin pratik eğitim olmasını şart koşmaktadır. Bu pratik eğitim sürecinde biçimlendirici (*formative*) ve düzey belirleyici (*summative*) geribildirimin bulunması gerektiği belirtilmektedir (O'Connor, 2018). Bunun başarılması, yapılan pratik veya klinik uygulamanın gözlem sonucunda değerlendirilmesi ile mümkündür.

Değerlendirme amaçlarına göre biçimlendirici ve düzey belirleyici olarak sınıflandırılabilmektedir (Vendrely, 2002).

1.1.2. Biçimlendirici-*formative* (eğitim durumunda yapılan) değerlendirme

Öğrencilerin öğrenim sırasındaki öğrenmesini izlemek, karşılaştığı güçlükleri ortaya çıkarmak ve gerekli değişiklik veya düzeltmeleri yapmak için yapılan bir değerlendirmedir. Eğitim süresince devam eden bir değerlendirme ve geribildirim işleyişini ifade eder. Öğretim elemanı ile öğrencinin öğrenimi geliştirmeye yönelik sürekli düzenleme veya değişiklik yapma imkânı sağlar. Temel amaç öğrenmenin geliştirilmesidir. Genellikle, öğrenim anındaki bilgi ile sınırlıdır.

Kısa ödevler, bilgi sınamaları (quizler), Sokratik sorgulamalar veya öğreticinin gözlemleri örnek olarak verilebilir. Vize veya düzey belirleyici sınavlar öncesinde öğrencinin bilgi ve becerisinin pekiştirilmesi veya düzeltilmesi amaçlanır. Fizyoterapi ve rehabilitasyon öğretiminde dersliklerde, laboratuvarlarda (anatomi, elektroterapi vb), pratik uygulama salonlarında (temel ölçme ve değerlendirme, tedavi hareketleri, egzersiz eğitimi, vb) veya kliniklerde yapılan dönem içi değerlendirmeler örnek olarak verilebilir.

Biçimlendirici değerlendirme, düzey belirleyici değerlendirmeye aşırı bağımlılığı ortadan kaldırmak için eğitimde desteklenmektedir. Sağlık alanında giderek önem kazanmaktadır. Ancak, biçimlendirici değerlendirmenin geri bildirim ve yansıtma süreçlerini de kullandığı düşünüldüğünde, daha fazla yer verilmesi ve düzey belirleyici değerlendirmeye olan oranının artırılması gerektiği savunulmaktadır (Chong, 2020).

1.1.3. Düzey belirleyici-*summative* (Programın sonunda yapılan) değerlendirme

Düzey belirleyici değerlendirmede temel hedef öğrencinin pratik veya teorik başarısının belirli norm veya ölçütlerle değerlendirilmesidir. Öğretilen bilgi paketinin veya modülün sonunda, vize dönemlerinde veya dönem sonlarında gerçekleştirilir. Öğrencilerin başarılı veya yetkin olup olamadıklarına yönelik karar verebilme sürecini içerir. Biçimlendirici değerlendirmede ‘değerlendirme’ bir süreçken, düzey belirleyici değerlendirmede ‘değerlendirme’ bir çıktıdır. Biçimlendirici değerlendirmede öğrenim sürecinin gözlenmesi hedeflenirken düzey belirleyici değerlendirmede puanlama hedeflenir (O’Connor ve diğerleri, 2018).

Usherwood ve diğerleri, 1995 yılında yaptıkları çalışmada tıp fakültesi birinci sınıf öğrencilerinde yeterlik temelli yaklaşımı ekleyerek düzey belirleyici değerlendirme geliştirmişlerdir. Başlangıç olarak, öğrencinin öğrenme süreçlerinin araştırılmasını hedefleyen çalışmanın sonucunda, biçimleyici değerlendirmelerin, düzey belirleyici değerlendirmelerle birlikte bir bütün olduğunu göstermişlerdir. Araştırmacılar ayrıca, sağlık eğitiminde öğrenim süreçlerinin, öğrenim çıktıları kadar önemli olduğunu vurgulamışlardır (Usherwood ve diğerleri, 1995).

Eğitimde kullanılan ölçme ve değerlendirme yöntemleri geleneksel ve alternatif (tamamlayıcı) yöntemler olarak iki ana grupta açıklanabilir.

1.1.4. Geleneksel ölçme değerlendirme yöntemleri

Geleneksel ölçme değerlendirme yöntemleri, genel olarak kâğıt-kalem kullanılarak yapılan uygulamaları tanımlar. Fizyoterapide iki türü daha çok kullanılmaktadır: 1) Cevap seçme (*selected response*): Örneğin, çoktan seçme, eşleştirme veya doğru/yanlış. 2) Cevap verme/cevap sağlama (*supply response*): Örneğin, kısa cevaplar, ödev veya denemeler. Her iki yöntemde de teorik bilginin veya pratik bir uygulamanın teorik olarak ifade edilebilmesi değerlendirilebilir. Cevap seçmede, cevaplamanın ve cevabın kısa sürede okunması olumlu yönlerdendir. Öğrenciyi tek bir cevapla sınırlaması, cevabı tahmin etmeye yönlendirmesi ise olumsuz yönleridir. Cevap vermede, daha derinlemesine bir değerlendirme ve araştırma gerektirmesi olumlu yönüdür. Cevaplanması ve cevapların değerlendirilmesi daha çok zaman alır. Geleneksel değerlendirmede, değerlendirme süresinin kısalığı olumlu yönü olarak gösterilebilir. Her iki geleneksel değerlendirme, önceki bilgi veya becerinin ölçülmesinde kullanılır. Bu tür değerlendirmeler, sağlık alanında öğrenim gören öğrencilerde daha çok, *board* (bir kurul tarafından değerlendirilen) sınavlarına (TUS, DUS, vb) veya genel lisans bitirme sınavlarına hazırlıkta önemli bir pratik deneyim sağlayabilir (Usherwood ve diğerleri, 1995).

Geleneksel ölçme değerlendirme, vaka temelli problem çözme ve karmaşık beceri performansının ölçülmesinde sınırlı bilgi verir. Fizyoterapi ve rehabilitasyon eğitimi, bilgi toplama, problem çözme, karar verme, karmaşık beceri performansı gibi kavramları içerir. Bu temel kavramlar, klinik çalışma ve mesleki uygulamalara başlamada kritik öneme sahiptir. Bu

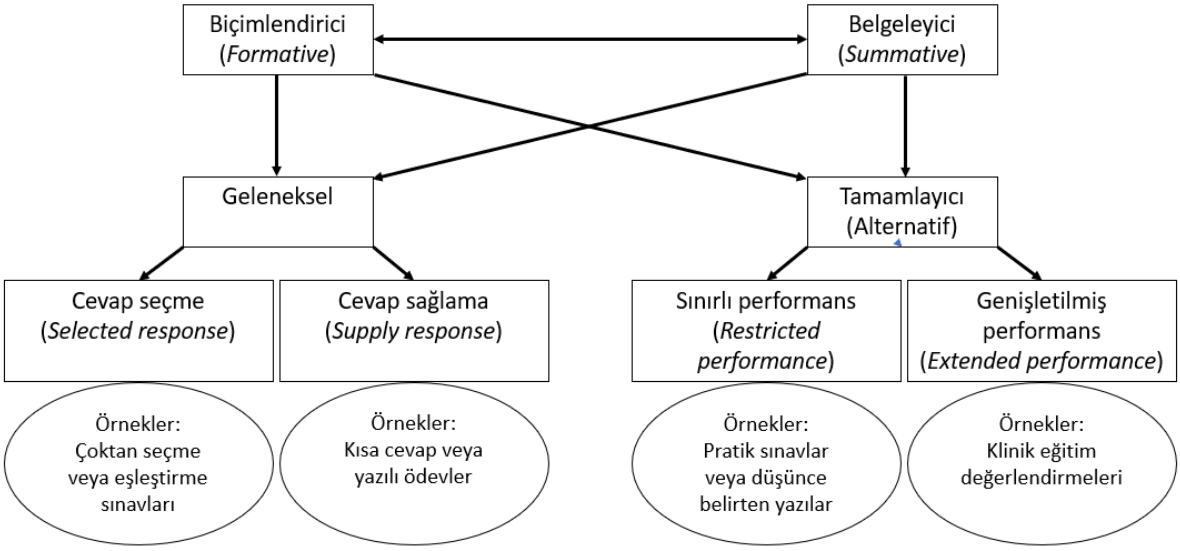
nedenle fizyoterapi ve rehabilitasyon öğretim elemanları, geleneksel değerlendirme yanında değerlendirilecek karmaşık beceriye ve problem çözmeye özel, diğer alternatif değerlendirme yöntemlerini de kullanmalıdırlar (McGinty, 2000).

1.1.5. Tamamlayıcı ölçme değerlendirme yöntemleri

Tamamlayıcı ölçme değerlendirme, kâğıt kalem testleri dışındaki yöntemler olarak tanımlanmaktadır. Tamamlayıcı değerlendirme, "alternatif" kelimesinin yerine tercih edilmektedir. Alternatif kelimesi, bir şeye karşı olma durumu değil, farklı bir seçeneğe karşılık gelmek anlamı içermektedir. Başka bir ifadeyle, kelime anlamı ile diğer yöntemlere karşı bir alternatif sunmak değil, diğer tüm değerlendirme yöntemlerin de dahil edildiği alternatif oluşturma süreçleri şeklinde kabul edilmelidir (Kutlu, Doğan ve Karakaya, 2017).

1.1.5.1 Fizyoterapi ve rehabilitasyonda tamamlayıcı değerlendirme

Karmaşık beceri performansının veya problem çözme yeteneklerinin, daha önce gösterildiği şekilde göstermesi veya oluşturması istenir. Öğrencilere öğrenmek istenilen alanda bir görev verip, o görevdeki performansının değerlendirilmesi hedeflenir. Tamamlayıcı değerlendirmeler, geleneksel yönteme göre, becerilerin veya verilen karmaşık görevlerin daha gerçekçi bir şekilde değerlendirilmesini amaçlayacak şekilde oluşturulur. Öğrenilen temel bilgi ile bilginin pratikte veya klinikte uygulaması arasında köprü oluşturur. Fizyoterapi ve rehabilitasyon programlarında, pratik sınavlar, klinik değerlendirmeler ve ödevler örnek olarak verilebilir (Vendrely, 2002).



Şekil 1: Değerlendirmeye Şematik Bakış (Vendrely, 2002)

(*Yazardan (Ann Vendrely) 6 Ocak 2021 tarihinde izin alınmıştır.*)

Sınırlı performans ve genişletilmiş performanslar tamamlayıcı değerlendirme türleridir. Sınırlı performansa örnek olarak, rol oynama veya kısa pratik değerlendirme verilebilir. Bir vaka senaryosu üzerinden yapılan değerlendirme veya fizyoterapiye özgü ölçme veya uygulama yöntemlerinden bir bölümünün belirli bir zaman içerisinde performansa dönüştürülerek değerlendirilmesidir. Geleneksel değerlendirmeye göre pratik uygulamaya yönelik daha gerçekçi ve kapsamlı bilgi verir. Fizyoterapi ve rehabilitasyon alanında sıklıkla kullanılan değerlendirme türüdür. Klinik alanda staj veya gerçek vaka üzerinde değerlendirme yapmadan önce, temel bilgi ve becerilerin değerlendirilmesine güvenli bir şekilde izin vermesi en önemli olumlu yanıdır (Venderly, 2002).

Fizyoterapi ve rehabilitasyon eğitiminin ilk iki yılında öğrencinin öğrenmesi gereken uygulamalı fizyoterapi becerileri pratik ve teorik ders kapsamında verilir. Hasta değerlendirmeye yönelik kas testi, postür analizi, eklem hareket açıklığının ölçülmesi kas kısalık testleri ve tedaviye yönelik egzersiz uygulamaları, yumuşak doku mobilizasyonu ve elektroterapi becerileri örnek olarak verilebilir. Fizyoterapi ve rehabilitasyona özgü bu becerilerin tedavinin başarısında doğrudan etkilidir. Doğru yapılmadığında veya yetersiz yapıldığında, en iyi ihtimalle etkisiz ve en kötü ihtimalle zararlıdırlar (Holey, 1993).

Öğrenimin ilk yıllarında yapılan geleneksel bu tür performans değerlendirmeleri, becerileri teknik düzeyde test eden ve gerçek hasta üzerinde herhangi bir değerlendirme veya

tedavi uygulaması yapılmadan önce gerçekleştirilen sınavlardır. Bir öğrenci tarafından yürütülen bir tedavinin etkinliği, klinik deneyim başlamadan gerçekçi olarak değerlendirilemez. Bir fizyoterapi ve rehabilitasyon yöntemi olarak hasta için yapılan değerlendirme veya tedavi uygulamalarının etkinliği veya doğru yapılıp yapılmadığı ancak hasta üzerinde değerlendirilebilir. Bu nedenle, bu aşamada öğrenciden beklenen, tedavinin etkinliğinden çok, hasta değerlendirme ve tedavisinde uygulanacak olan tekniklerdeki yeterliliktir (Vendrely, 2002).

Yeterliliğin tek bir bilgi, beceri veya mesleki değer olduğunu düşünürsek, yetkinliği bir yeterlilikler bütünü olarak kabul edebiliriz (Medley, 1984). Dolayısıyla, yetkinlik yeterlilikten sonra gelen bir uygulamadır. Bununla birlikte, sınav değerlendirmesi, öğrencilerin performansına ilişkin yargıda bulunmayı gerektirir. Bu, doğrudan hasta üzerinde belli derecede baskı altında çalışırken entelektüel ve psikososyal yeterlilikler yanında, üst düzey teknik yeterlilik göstermelerinin beklendiği klinik uygulama ortamından farklıdır. Öğrenci, klinik öncesi ağırlıklı olarak uygulama teknikleri üzerinde çalışmaktadır. Bu aşamada öğrenciden beklenen, beceriyi üretmek için bir araya gelen unsurları bulması ve güvenli bir şekilde beceri edinimini sağlamasıdır (Miller, 1990).

Fizyoterapide genişletilmiş performans değerlendirmesi ise, klinik eğitim pratiği sırasında yapılan performans değerlendirmesidir. Sınırlı performansa göre daha kapsamlıdır. Ancak değerlendirme, daha fazla zaman ve profesyonel beceri gerektirir (Vendrely, 2002). Klinik performans değerlendirme araçları olarak geliştirilen ve klinik ortamda öğrenci performansını değerlendirmeyi amaçlayan araçlar bulunmaktadır. “Assessment of Physiotherapy Practice (APP)” (Dalton ve diğerleri, 2011), “Blue MACS” (Hrachovy ve diğerleri, 2000) ve “PT CPI” (2006 versiyonu) (Roach ve diğerleri, 2006) örnek olarak sayılabilir.

Fizyoterapi ve rehabilitasyon lisans öğrencilerinden, tıp öğrencilerinin aksine, mezuniyetlerinden hemen sonra doğrudan bağımsız olarak çalışmaları beklenir. Benzer durum hemşireler için de geçerlidir (O’Connor vd., 2018). Bu nedenle performans değerlendirmesinin klinik öncesi ve klinik uygulamalarda değerlendirilmesi, gerekli düzenleme ve değişikliklerin öğrenciye göre adapte edilmesi önemlidir.

1.1.6. Objektif Yapılandırılmış Klinik Değerlendirme (*Objective Structured Clinical Examination, OSCE*)

İlk kez 1975 yılında tıp eğitiminde Harden ve diğerleri tarafından ortaya konmuştur (Harden vd, 1975). Birçok sağlık eğitimi alanında klinik becerilerin değerlendirilmesinde kullanılmaya başlanmıştır. Birden çok istasyondan oluşan bu yöntemde farklı klinik performans değerlendirilmektedir. İçeriğinde gözlem, görüşmeler ve gözlem dosyaları bulunabilmektedir (Denat ve Tuğrul, 2012; Başaranoğlu, 2018).

OSCE klinik ortam oluşturularak yapılan bir sınav türüdür. İstenen performansın izlenmesi üzerine kurulmuştur. OSCE, istasyonların kurulması, teknik ekipman gerekliliği, uygun fiziksel alanların oluşturulması gibi nedenlerle kurulması zor, zaman alan ve ekonomik olmayan sınavlardır. Öğrenci sayısının fazla olması uygulanabilirliğini azaltan bir diğer etkendir. Artan öğrenci sayısıyla birlikte puanlayıcının yorgunluğunun süreci etkilediğine dair çalışma mevcuttur (Haider vd., 2018). Fizyoterapi alanında karşılaşılan bir diğer sorun da uygulama becerilerinin değerlendirilmesinden daha çok, görüşmeye dayalı değerlendirmeler için daha uygun olmasıdır. Değerlendirmeye çalıştığı klinik uygulamaları parçalar halinde yapması, bu parçaların bir bütünün tamamını temsil edemeyeceği ve genel bakış açısı oluşturması adına yetersiz kalabileceği vurgulanmaktadır (Daniels ve Pugh, 2018).

1.1.7. Objektif Yapılandırılmış Pratik Değerlendirme (*Objective Structured Practical Examination, OSPE)*

Klinik ortamda pratik gözlem ve uygulama aşamasına gelmeyen öğrencilerde (ilk 2 yıl içerisinde) klinik değerlendirme yerine, uygulamanın pratik olarak (bir başka öğrenci veya eğitici üzerinde) değerlendirilmesini amaçlayan ve Yapılandırılmış Objektif Klinik Sınav yönteminden uyarlanan Yapılandırılmış Objektif Pratik Sınav (OSPE) geliştirilmiştir (Olivier, vd., 2013). İlk kez Nayar ve diğerleri (1986) tarafından geliştirilen bu yöntem fizyoloji eğitiminin değerlendirilmesinde kullanılmıştır. Anatomi ve fizyoloji gibi temel sağlık eğitiminde yer alan derslerin değerlendirilmesinde daha çok kullanılmıştır (Yaqinuddin vd, 2013).

Son yıllarda yapılan çalışmalar, Yapılandırılmış Objektif Klinik/Pratik Sınav'ın, klasik olarak yapılmakta olan pratik sınavlara olan üstünlüklerine rağmen, öğrenci üzerinde stres oluşturduğu, maliyetinin fazla olduğu, sınav süresinin uzunluğu, sınav yapanın yorgunluğunun puanlamada etkili olduğu, tecrübeli öğretim elemanlarının bu yöntemi uzun ve aşırı ayrıntılı bulması, az sayıda öğrencinin değerlendirilebilmesi gibi dezavantajları olduğunu göstermiştir (Ribeiro vd., 2019)

1.1.8. Fizyoterapide Diğer Klinik Performans Değerlendirme Araçları

Fizyoterapi alanında klinikte hasta ile karşılaşan öğrencilerin performansının değerlendirilmesine yönelik geliştirilen diğer performans göstergelerinin büyük bölümü, öğrencinin klinik ortamda gözlenmesine dayalı değerlendirme araçlarıdır. Bu araçlardan özellikle "Physical Therapy Clinical Performance Instrument (PT CPI)", "Assessment of Physiotherapy Practice (APP)", "Clinical Performance Assessment Form, Physical Therapy Manual for the Assessment of Clinical Skills" örnek olarak sayılabilir. Bunlardan APP ve PT CPI en çok geçerlik ve güvenirlik çalışması yapılan araçlardandır (O'Connor vd., 2018).

1.1.9. Performansa Dayalı Durum Belirleme

Performans, "bir öğrenme görevine yönelik çabalar ve ortaya konulan ürün" (Büyüköztürk, 2007), performans değerlendirme ise "gözlem ve kanıya dayanan değerlendirme" (Palm, 2008) olarak ifade edilmektedir (Büyükkıdık ve Anıl, 2015). Bir diğer ifade ile, özel kavramları, bilgileri veya becerileri gerçek durum veya şartlarda gösterme yöntemi olarak da belirtilebilir. Performansın ölçülmesinde öğrenciler, kendilerine verilen performans görevlerini yerine getirebilmek için ön bilgilerini sorgular ve elde ettiği bilgileri kullanır (Bekiroğlu, 2008; Yılmaz Nalbantoğlu ve Gelbal, 2011). Performansın ölçülmesiyle, öğrencilerin öğrenmeleri pekiştirilmiş, bilgiyi nasıl anladıkları ve bu bilgiyi nasıl kullandıkları belirlenmiş olur (Brualdi, 1998)

1.1.10. Sağlık alanında performans değerlendirme

Miller'ın Klinik Değerlendirme için Ustalık Piramidi:

Miller, karmaşık beceri gerektiren durumlarda tatmin edici bir karar verebilmek için tek bir değerlendirmenin yeterli olamayacağını belirtmiştir. Sağlık eğitiminde dört aşamalı bir

yapı önermiştir. Bu yapının temelini bilgi oluşturmaktadır ve çoktan seçme gibi standart objektif yöntemlerle test edilebilir. İkinci aşamada yeterlik yer alır. Bu aşama öğrenci vaka değerlendirmesinde olduğu gibi, bilgisini nasıl kullanacağını bilir. Üçüncü aşama ise öğrencinin performansını değerlendirir. Bu aşamada öğrenciden belirli bir işlemi göstermesi beklenir. Son aşama ise, edinilen becerileri klinik uygulamada bağımsız olarak seçebilmesi ve uygun bir şekilde kullanabilmesidir.

Fizyoterapi ve rehabilitasyon eğitiminde tek bir performans değerlendirme yöntemini önermek mümkün değildir. Performans yöntemlerinin tümünde güçlü ve zayıf yönler bulunmaktadır. Özellikle, güvenirlik ve geçerlik yönünden değişen oranlarda sonuçlar bildirilmektedir. Fizyoterapide yalnızca uygulama yöntemleri alanında değil, fizyoterapinin farklı alanlarında da uygulama farklılıkları mevcuttur (kardiyopulmoner, pediatrik, ortopedik veya nörolojik fizyoterapi ve rehabilitasyon gibi). Bu nedenle değerlendirme süreçlerinin ve yöntemlerinin belirlenmesinde, akademik ekibin tüm koşulları göz önünde bulundurarak dikkatli bir şekilde planlama yapması gerekir (Sattelmayer vd, 2017; Sakurai vd, 2014).

1.1.11. Dereceleme Ölçeği ve Dereceli Puanlama Anahtarları

Dereceli puanlama anahtarı (rubrik), öğrencinin durumunun saptanmasında ve takibinde uygulanan puanlama rehberi olarak tanımlamıştır. Bir çalışma parçası için belirlenen ölçütleri veya neyin önemli olduğunu listeleyen ve her bir kriter için aşamalı puanlamayı (örneğin, zayıf, orta, iyi gibi) ifade eden bir değerlendirme aracıdır. Değerlendirilen her çalışma için, ölçütler belirtilmekte ve nelerin yapılacağı sıralı olarak gösterilmektedir Goodrich, 1997; Kutlu vd, 2017). Değerlendirme sonucunda öğrenci ve öğretmen gelişim düzeyini izleyebilmektedir. Dereceleme ölçeğine ise değerlendirilen her madde için performansın kalitesi yok, zayıf, orta, yeterli gibi sıra sayılarıyla belirtilir (Turgut ve Baykul, 2019). Başka bir ifade ile, hangi performansın hangi seviyede yapıldığını gösterir. Dereceli puanlama anahtarında ise, performansın nasıl/ne kadar yapıldığında karşılığında ne kadar puan alacağını belirten puanlama anahtarının olmasıdır (Yıldırım Hacıibrahimoğlu, 2018).

Dereceli puanlama anahtarları a) Analitik dereceli puanlama ve b) Bütünsel dereceli puanlama anahtarı olmak üzere 2'ye ayrılırlar. Her iki yöntemin Genel ve Göreve özel alt başlıkları bulunmaktadır.

Bütünsel Dereceli Puanlama Anahtarı:

Bütünsel dereceli puanlama anahtarında, öğrencinin performansının tamamına tek bir puan verilmektedir. Performans genel olarak değerlendirilmekte ve ayrıntılı analiz edilememektedir. Puanlama yapılacak performansın alt görevleri birbirinden ayrı değerlendirilemeyeceği durumlarda tercih edilmektedir. Farklı görevlerde farklı performans gösteren öğrencilerin aynı puanı alabilmesi, görev basamaklarında öğretim sürecindeki eksiklik ve geliştirilmesi gereken alanlara yönelik ayrıntı vermemesi olumsuz yönlerindendir. "Genel izlenim"e dayalı bütünsel değerlendirmede bir performans baştan sona izlenerek edinilen genel izlenim sonucunda puanlama yapılmakta ve genel olarak bir toplam puan verilmektedir (Kutlu, Doğan ve Karakaya, 2017).

Analitik Dereceli Puanlama Anahtarı:

Analitik dereceli puanlama anahtarında, performans değerlendirmesinde performansı oluşturan maddeler görev tanımıyla belirlenir ve görevlere dereceli puanlamalar verilir. Her bir görev için öğrenci ve tüm grup düzeyinde geliştirilmesi gereken konuların belirlenmesi ve değerlendirilmesi mümkündür. Bütünsel yaklaşıma göre olumsuz yönü hazırlanma ve puanlanmasının daha fazla zaman almasıdır.

Analitik dereceli puanlama anahtarında ölçülmesi hedeflenen performans alt boyutlara (görevlere) ayrılır. Görevler için öğrencinin gösterebileceği performans seviyeleri belirlenir. Puanlama anahtarının öğrenci performansını tümüyle yansıtıp yansıtmadığına dikkat edilmelidir. Kullanılan ifadelerin açık ve anlaşılır olmasına özen gösterilmelidir (Kutlu, Doğan ve Karakaya, 2017).

"Dereceleme ölçeği"nde, puanlayıcının puanlama yapılacak özelliği iyi bilmesi gerekmektedir. Ölçülecek performans konusunda ön bilgiye sahip olması önemlidir. Puanlayıcının görev için istekli ve yeterli zaman ayırması beklenmektedir (Deliceoğlu, 2009; Yıldırım Hacıibrahimoğlu, 2018).

1.2. Araştırmanın Amacı ve Önemi

Bu çalışmada, fizyoterapi ve rehabilitasyon öğrencilerinde eklem hareket ölçüm becerilerinin pratik değerlendirilmesinin genellenebilirlik kuramına ve klasik test kuramına göre incelenmesi amaçlanmıştır.

Bu kapsamda iki farklı puanlama yöntemi olan bütünsel puanlama ve analitik puanlama yöntemleri için puanlayıcılar arasındaki tutarlılığa bakılacak; iki puanlama yöntemiyle elde edilen puanların güvenirliği genellenebilirlik ve klasik test kuramına göre incelenecek ve sonuçları betimsel olarak karşılaştırılacak; genellenebilirlik kuramına göre elde edilen sonuçlardan sonra karar çalışması yapılacaktır. Puanlayıcılar arası güvenirlik ve ilişki çalışmalarının Klasik Test Kuramına göre incelenmesi amaçlanmıştır. Ayrıca, analitik puanlama türünde Klasik Test ve Genellenebilirlik Kuramlarının bütünsel (genel izlenim) ve dereceleme ölçeği kullanılarak elde edilen ölçümlere ilişkin güvenirlik belirlemedeki yaklaşımlarının incelenmesi hedeflenmiştir.

Çalışmada, analitik puanlama kapsamında "dereceleme ölçeği" ve bütünsel puanlamada da "genel izlenim" kullanılmıştır. Dereceleme ölçeğinde her madde için puanların karşılıklarında dereceleme bulunmakta; ancak, her bir derece için hangi durum veya durumlarda hangi derecenin verileceği ayrıntılı olarak belirtilmemektedir. Genel izlenimde de tüm performans izlendikten sonra değerlendirici, genel olarak performansa karşılık tek bir puan vermektedir.

Fizyoterapi ve rehabilitasyon alanında, performansın Genellenebilirlik Kuramına göre değerlendirilmesinde yapılan ilk çalışma olması nedeniyle, sonraki çalışmalarda fizyoterapi ve rehabilitasyon eğitiminin standardize edilmesinde çok önemli bir yer tutan eğitimde ölçme ve değerlendirme için farklı bir başlangıç oluşturması diğer bir amaç olarak planlanmıştır. Mesleki eğitim standartlarının geliştirilmesi ve ileriye taşınmasında uluslararası platformda katkı sağlaması hedeflenmektedir. Eklem hareket sınırının ölçülmesi gibi bilgi ve uygulama farklılıklarının bireyden bireye değişeceği değerlendirmelerde, puanlayıcıların öğrencilere verdikleri puanlar arasındaki ilişki önemlidir. Kullanılacak puanlama anahtarları analitik, dereceli, bütünsel veya genel izlenim olsun, değerlendiriciye ait yargılar da etkili olmaktadır (Erkuş vd, 2017; Yıldırım Seheryeli, 2018).

1.3. Problem Cümlesi

Fizyoterapi ve rehabilitasyon eğitiminde, öğrencilerin mesleki temel pratik uygulamalı derslerinden olan Temel Ölçme ve Değerlendirme dersi kapsamında performans değerlendirmesi dereceleme ölçeği ve genel izlenim kullanılarak yapılan gonyomektik ölçümün güvenirlikleri Klasik Test Kuramı ve Genellenebilirlik Kuramına göre farklılık gösterir mi?

1.3.1. Alt Problemler

1. Dereceleme Ölçeği ve Genel İzlenim kullanılarak elde edilen ölçümler puanlayıcıların test ve tekrar test puanlamalarına göre farklılık göstermekte midir?
2. Farklı puanlayıcıların Dereceleme Ölçeği ve Genel İzlenim ile yaptıkları puanlamalar, Sınıf İçi Korelasyon Katsayısı ve Pearson Korelasyon katsayısına göre farklılıklar göstermekte midir?
3. Genellenebilirlik Kuramına göre, dereceleme ölçeği için yapılan öğrenci x öğretim üyesi x görev (birey x puanlayıcı x madde, b x p x b) çapraz deseninde nasıl sonuçlar vermektedir?
4. Genellenebilirlik Kuramı ile elde edilen sonuçlara göre yapılacak Karar çalışması nasıl farklılıklar göstermektedir?
5. Klasik Test ve Genellenebilirlik kuramlarının birbirlerine göre üstünlükleri var mıdır ve ne tür farklı sonuçlara ulaşmamızı sağlamaktadırlar?

1.4. Sayıltılar

1. Oluşturulan dereceleme ölçeği formunun hazırlanmasına katkıda bulunan alanında uzman ve tecrübeli öğretim elemanları, görüşmelere ciddiyetle yaklaşmışlar ve görüşlerini uluslararası örneklerle desteklemişlerdir.
2. Farklı dört üniversitede görev yapmakta olan puanlayıcılar her iki puanlama türlerine göre değerlendirmeleri birbirlerinden bağımsız olarak yaptıkları varsayılmıştır.
3. Performans değerlendirilmesi için alınan görüntülerin, yapılan performansın değerlendirilebilmesine tamamen izin verdiği varsayılmıştır.

1.5. Sınırlılıklar

1. Çalışma, gonyometrik ölçümde performansın değerlendirilmesi için seçilen yalnızca bir eklemdeki ölçümle sınırlıdır.

2. Çalışma pandemi öncesi sınıf ortamında planlanmıştır. Ancak pandemi sonrası yapıldığı için görüntü kaydıyla yapılabilmiştir.

1.6. Tanımlar

Bu bölümde, araştırmanın kuramsal temellerinde Klasik Test Kuramı ve Genellenebilirlik Kuramına değinilmiştir.

1.6.1. Klasik Test Kuramı

Klasik Test Kuramı (KTK), gözlenen özelliklerin gerçek değerini bulmayı amaçlar. Bu teoride, gözlenen puanlardan gerçek puanlar elde edilmeye çalışılır. Bu nedenle KTK'ya gerçek puan teorisi de denilmektedir (Baykul, 2015; Yıldıztekin, 2014). KTK, eğitim, psikoloji ve diğer performans değerlendirmelerinde güvenirliğin analizinde de kullanılmıştır. Bir ölçme teorisi olarak KTK'da amaç, değerlendirilen bir özelliğin gerçek değerini hesaplamaktır. KTK, ölçmeye karışan çeşitli hatalar yüzünden bu gerçek değerin, ölçme yoluyla doğrudan elde edilemeyeceğini, gözlenen puanlar yardımıyla kestirilmeye çalışılacağını belirtir (Baykul, 2015:91; Büyükkıdık, 2012:41). KTK'da, ölçüm hatasının rastlantısal olduğu ve bu hataya ait puanın, her ölçme için gerçek puandan bağımsız olduğu kabul edilmektedir. Hata puanının, bir başka ölçmedeki hata puanından bağımsız olduğu varsayılır. Bu çerçevede KTK'nin temel varsayımları aşağıdaki gibi açıklanabilir:

KTK'da temel denklemi Gözlenen Puan (X), Gerçek Puan (T) ve Hata (E) oluşmaktadır. Formül ile ifade edilecek olursa:

X = T + E'dir.

Gerçek puan hata puanı ile ilişkisiz kabul edilir. X=T+E formülü esas alınarak, Gözlenen puan varyansı aşağıdaki gibi gösterilir:

$\sigma^2_X = \sigma^2_T + \sigma^2_E$

(σ^2_X : Gözlenen puan varyansı, σ^2_T : Gerçek puan varyansı, σ^2_E : Hata puanı varyansı)

KTK'da güvenirlik katsayısı (ρ), gerçek puan varyansının (σ^2_T), gözlenen puan varyansına (σ^2_X) bölünmesiyle elde edilir:

$$\rho = \sigma^2_T / \sigma^2_X$$

KTK'nın varsayımları aşağıdaki gibi açıklanmaktadır:

- Hata puanlarının beklenen değeri sıfıra eşittir.
- Ölçme evreninde gerçek puanlarla hata puanları birbirinden bağımsızdır. Diğer bir ifadeyle, hata puanı ile gerçek puan arasındaki ilişki sıfıra eşittir.
- Her ölçümdeki hata puanı diğer bir ölçümdeki hata puanından bağımsızdır. Dolayısıyla, farklı ölçümlerdeki hata puanları arasında ilişki yoktur (Crocker ve Algina, 1986; Yıldıztekin, 2014).

Güvenirlik katsayısı, gözlenen puan varyansı veya temel denklem olan gözlenen puanın belirlenmesinde her klasik test modeli için tek bir hata puanı veya varyansı bulunmaktadır. Çoklu hata kaynaklarının açıklanmasında tek bir işlem veya yöntem yeterli değildir. Böyle durumlarda, çeşitli güvenirlik yöntemlerine gerek duyulmaktadır (test-tekrar test, içi tutarlılık, puanlayıcılar arası tutarlılık, paralel form, toplam-alt başlıklar korelasyonu, gibi) (Suen ve Lei, 2007:1; Yılmaz Nalbantoğlu, 2012:5; Yıldıztekin, 2014:12). Çoğu zaman, KTK'ya göre yapılan güvenirlik çalışmasında birden fazla yöntemin uygulaması gerekmektedir.

Güvenirlikte ölçme hataları, birbirinden bağımsız olarak yapılan ölçme sonuçlarından elde edilen değer ile ölçülen özelliğin gerçek değeri arasındaki fark olarak tanımlanmaktadır (Özçelik, 2013). Ölçmede hata kaynakları, ölçmeyi yapandan kaynaklı hata, kullanılan ölçme aracından kaynaklanan hata, yapılan ölçme ortamından kaynaklanan hata ve ölçülen kişiden kaynaklanan hatalar olarak sayılabilir. Hata türleri ise 3 başlıkta toplanabilir: 1) Sabit hata: Kaynak ve miktarın belirtilebildiği hata türüdür ve her bir ölçüm için değişiklik göstermez. 2) Sistematik hata: Ölçmeye karışan hata miktarı ölçümden ölçüme değişiklik gösterir. Hatanın kaynağı ve miktarı bellidir. 3) Tesadüfi hata: Geçerliliği doğrudan etkileyen hata türüdür. Ölçme sonuçlarına etki eden, aynı zamanda kaynağı, yönü ve ölçme sonuçlarını hangi yönde etkileyeceği kestirilemeyen hata türüdür (Köse, 2014; Yıldırım Seheryeli, 2018).

Birden fazla boyutu bulunan ölçümlerde, boyutlar arasındaki iç tutarlılık düşük olabilirken, test-tekrar test güvenirliği yüksek çıkabilir. Bu durum KTK için karşılaşılan kısıtlılık ve farklı yorumlara neden olabilecek çelişkili durumlara neden olabilmektedir (Atılgan H, 2019:3).

1.6.1.1. KTK'da güvenirliğin değerlendirilmesi

Yapılan herhangi bir ölçmede, farklı etkenler hataya neden olabilir. Bu çalışmanın temelinde, fizyoterapi ve rehabilitasyon eğitiminde önemli bir yer tutan pratik performansın değerlendirmesinde, öğrenci, diğer bir ifadeyle birey, değerlendiren öğretim elemanı (puanlayıcı), değerlendirilen birey sayısı, performansın ölçülmesi istenen eklem, üzerinde ölçüm yapılan bireyin durumu gibi pek çok etken hataya neden olabilir. Bu çalışmada, puanlayıcıların iki farklı puanlama türü (analitik ve bütünsel) ile aynı eklemde yapılan ve gonyometre ile değerlendirilen eklem hareket açıklığının ölçülmesinin değerlendirilmesinde puanlayıcılar arası etkiler belirlenmeye çalışılmıştır.

Test-tekrar test güvenirliği

Puanlayıcıların yaptığı değerlendirmede, aynı puanlayıcının farklı zamanlarda yaptığı puanlamaların arasındaki tutarlılığın (test-tekrar test güvenirliği) test edilmesidir. Puanlayıcıların aynı yöntemle yaptıkları puanlamalarda, elde edilen değerin kaynağının puanın değerinden mi, yoksa puanlayıcının etkisinden mi kaynaklandığı bilinemez. Böyle durumlarda, bazı araştırmacılar, bu problemin ortadan kaldırılmasında aynı puanlayıcının farklı zamanda aynı yöntemle tekrar değerlendirmesinin incelenmesini önermektedirler (Gwet, 2020: 207).

Test-tekrar test güvenirliğinde elde edilen puan türüne göre Kappa katsayısı veya Sınıfiçi Korelasyon Katsayısı (Intraclass Correlaiton Coefficient, ICC) kullanılmaktadır. Sınıfiçi Korelasyon Katsayısı Fleiss tarafından Pearson korelasyon analizinden geliştirilmiştir. Sınıfiçi Korelasyon Katsayısı (Intraclass Correlaiton Coefficient, ICC), %95 güven aralığına göre, 0,50'den küçük ise zayıf, 0,50-0,75 arasında ise orta, 0,75-0,90 arasında iyi ve 0,90'dan büyük değere sahipse mükemmel güvenilirliğin göstergesidir (Portney and Watkins, 2015: 588; Koo and Li, 2016:155).

Puanlayıcılar arası güvenirlik

Puanlayıcıların yaptığı değerlendirmelerin güvenirliğinin test edilmesinde Sınıfiçi Korelasyon Katsayısı (Intraclass Correlaiton Coefficient, ICC) kullanılmaktadır. İki veya daha fazla puanlayıcının aynı yöntemle ve aynı bireylerde yaptıkları ölçümlerin güvenirliği incelenir.

İç tutarlılık güvenirliği

Çoklu puanlanan maddelerin yer aldığı testlerde güvenirlik, Cronbach Alfa istatistiği ile belirlenir. İç tutarlılığın belirlenmesinde kullanılan Cronbach Alfa katsayısı için kaynaklar 0,70'i kabul edilebilir düzey olarak vermektedirler. 0,80 üzeri değerin tercih edilebilir olduğu, 0,90 ve üzerinin ise mükemmel olarak kabul edilmesi gerektiği belirtilmektedir (Cotina, 1993; Taber, 2018).

1.6.2. Genellenebilirlik Kuramı

Genellenebilirlik kuramı (GK), hem klasik test kuramı ve hem de varyans analizinin çok yönlü bir uzantısı olarak, çoklu hata kaynaklarının aynı anda ele alınabildiği bir model olarak düşünülebilir (Güler, Kaya Uyanık ve Taşdelen Teker, 2012). KTK'daki gerçek puan varyansının yerini GK'da evren puanı varyansı alır. Bu durumda, genellenebilirlik katsayısı, evren puanı varyansının gözlenen puan varyansına oranıdır. Gözlenen puanın varyansı ise, evren puanı varyansı ve göreceli hata varyansının toplamıdır (Brennan, 2001; Nalbantoğlu Yılmaz, 2012).

Genellenebilirlik kuramının ilk başlangıcı, 1960'lı yılların başlarında Chronbach, Rajartman ve Gleser tarafından yapılan çalışmalarla olmuştur. Brennan'ın 1983 ve 2001 yıllarındaki çalışmalarla son şeklini almıştır (Brennan, 2001; Güler, Kaya Uyanık ve Taşdelen Teker, 2012). Bu kuram temelde varyans analizi üzerine kuruludur. Farklı desenlerde, desenlere göre elde edilen değişkenlerin her birine ve bileşenlerine ait varyans ve varyans yüzdeleri hesaplanmaktadır. Bu çalışmada kullanılan öğrenci (birey), öğretim üyesi (Puanlayıcı) ve analitik performans ölçümü için hazırlanan maddelere göre (madde) Genellenebilirlik çalışmasında aşağıdaki gibidir.

Tablo 1: Tamamen çaprazlanmış b x p x m desenli Genellenebilirlik çalışmasında değişkenlik kaynakları

Değişkenlik kaynağı	**Değişken türü**
Birey (öğrenci) (b)	Evren puanı (ölçme objesi)
Puanlayıcı (p)	Puanlayıcıların sıkılığının neden olduğu bütün bireyler üzerindeki sabit etki
Madde (m)	Bir maddeden diğerine değişen birey davranışlarından kaynaklı bütün bireyler üzerindeki sabit etki
b x p	Bireyin puanlanmasında puanlayıcılar arası tutarsızlık
b x m	Bir maddeden diğerine bireyin cevaplamalarındaki farklılık
p x m	Bir maddeden diğerine puanlayıcı sıkılığı arasındaki farkın neden olduğu sabit etki
b x p x m, e	Artık varyans

1.6.2.1. Çaprazlanmış ve yuvalanmış desenler

Bu iki kavramı çalışmamızdaki örnek üzerinden (performans değerlendirme) açıklayabiliriz. Bir birey (öğrenci), bir performans görevini (madde) yaparken, puanlayıcı tarafından değerlendirilmektedir (hazırlanan dereceleme ölçeği ile). Çalışmada, tüm öğrenciler, tüm puanlayıcılar tarafından, belirlenen dereceleme ölçeği ile değerlendirilmiştir. Bu uygulamanın deseni: b x p x m şeklindedir.

Tamamı çaprazlanmış desen çalışmaları daha çok tercih edilmektedir. Bu sayede, değişkenlik kaynaklarının oluşturduğu tüm yüzeylerden ve bunların birbirleriyle olan etkileşimlerinden kaynaklanan hataların kestirilmesine olanak verir. Bu desen aynı zamanda Karar çalışmaları için çok elverişlidir. Çaprazlanmış desenden farklı olarak yuvalanmış

desende, farklı öğrencilere farklı maddeler sunulur ve farklı puanlayıcılar bu farklı maddeleri cevaplayan farklı öğrencileri puanlar. Bu tümüyle yuvalanmış desendir. p : m : b olarak belirtilir. Desen çalışmaları, yalnızca çaprazlanmış veya yalnızca yuvalanmış değil, ikisinin bir arada olduğu desenler şeklinde de olabilir (Güler, Kaya Uyanık ve Taşdelen Teker, 2012).

1.6.2.2. Genellenebilirlik kuramında Karar çalışmaları

Genellenebilirlik kuramında yapılan Karar çalışmalarında, belirli bir amaç doğrultusunda karar vermek için Genellenebilirlik sonuçları ve bu sonuçlardan elde edilen analizlerden yararlanılır. Bu sayede, hata varyansının gerçek puan üzerindeki etkisinin azaltılmasına yönelik farklı durumlarda tekrar araştırma ve değerlendirme yapılması mümkün olur. Karar çalışması, kabul edilebilir hata varyansı ile farklı durum senaryolarına göre güvenirlik katsayısı kestirimleri yapmak amacıyla kullanılır (Shavelson and Webb, 2006; Nalbantoğlu Yılmaz, 2012; Güler, Kaya Uyanık ve Taşdelen Teker, 2012; Yıldıztekin, 2014).

Karar çalışması, iyi belirlenmiş ölçme süreçleriyle, kararlar alabilmek adına, varyans bileşenlerinin kestirimleri, kullanılması ve yorumlanması olarak düşünülebilir (Yıldıztekin, 2014:23; Brennan, 2001). Örneğin, bir Karar çalışmasında, bir puanlayıcı ekleyerek tahminin doğruluğunu ne kadar artırabildiğimizi görmek mümkündür. Güvenilirlikte bileşenlerin etkisi hakkında tahminler alabilmek, klasik test kuramına göre büyük bir avantajdır. Bir araştırmacı, bir Karar çalışması kullanılarak, değişkenlerin seviyeleri değiştirebilir ve yeni güvenilirlik tahminleri elde edebilir (Prion vd., 2016)

GK'nın avantajları şu şekilde özetlenebilir (Brennan, 2001; Atılgan, 2005):

GK ölçme durumunda bütün olası hata kaynaklarını birlikte ve eş zamanlı değerlendirmektedir. Bu sayede, KTK'da önemli bir sorun olan tek bir hata kaynağına yönelik birden fazla değerlendirme yöntemlerinin uygulanmasına gerek kalmamaktadır.

GK ile yapılan değerlendirmede mutlak ve göreceli değerlendirmeler için katsayı belirlenebilmektedir. KTK'da ise göreceli değerlendirme için güvenirlik hesaplaması söz konusudur. Mutlak değerlendirme için katsayılar her amaç için farklı katsayılılarla açıklanmaktadır.

GK'da yalnızca değerlendirilen değişken kaynakları değil, değişkenlik kaynaklarının farklı bileşenleri de aynı anda incelenebilmektedir.

GK içerisindeki Karar çalışmaları değişik senaryolarla, değişken kaynaklarının en uygun sayılarının belirlenebilmesine imkân verir.

GK'da, güvenirlik ve geçerlik için birbirinden farklı yöntemlerin ayrı ayrı belirlenmesine ve uygulanmasına gerek kalmayabilir. Burada, alınan örneklemin evrene genellenebilirliği test edildiğinden, geçerliliğinin de incelenmiş olduğu varsayılabilir.

1.6.3. Klasik Test Kuramı (KTK) ve Genellenebilirlik Kuramı (GK) karşılaştırması

KTK'da güvenirliğin hesaplanma yöntemleri kullanılan güvenirlik yöntemine göre farklılık gösterir. İki farklı zamanda gerçekleştirilen test-tekrar test güvenirliğinde olası en büyük hata, zamanın kendisi olarak dikkate alınır. Paralel formlar yönteminde ise temel hata kaynağı formlar olarak kabul edilir. İç tutarlılıkta ise (Örneğin Cronbach alfa) maddeler veya alt madde grupları hata kaynağı olarak incelenir. Bu farklı uygulama zorunlulukları, aynı zamanda KTK'nın kendi içerisindeki zorluk ve çelişkiyi de göstermektedir (Yelboğa ve Tavşancıl, 2010; Cronbach, Rajaratnam ve Gleser, 1963).

Klasik test kuramı soruları ve Genellenebilirlik kuramları karşılıkları aşağıdaki tablo halinde gösterilebilir (Prion, Gilbert ve Haerling, 2016; Bloch ve Norman, 2012):

Tablo 2: Klasik test kuramı soruları ve Genellenebilirlik kuramı karşılıkları.

Klasik test kuramı	Genellenebilirlik kuramı
• Bu sınavın puanlayıcı içi güvenilirliği nedir? • Bu sınavın puanlayıcılar arası güvenilirliği nedir? • Bu sınavın test-tekrar test güvenirliği nedir? • Bu sınavın puanlayıcılar arası güvenirliği- test-tekrar test güvenilirliği nedir? (Hesaplanamaz) • *Soru karşılığı yok.*	• Puanlayıcı içinde bu sınav puanlarını ne ölçüde genelleyebiliriz? • Puanlayıcılar arasında bu sınav puanlarını ne ölçüde genelleyebiliriz? • Sınav puanlarını duruma göre ne ölçüde genelleyebiliriz? • Puanlayıcılar arasında ve duruma göre sınav puanlarını ne ölçüde genelleyebiliriz? • … hata kaynağına göre sınav puanlarını ne ölçüde genelleyebiliriz? (…: potansiyel hata kaynağı)

İKİNCİ BÖLÜM

KAVRAMSAL ÇERÇEVE

2.1 Eğitim alanında yapılan çalışmalar

Doğan ve Anadol (2017), çalışmalarında tümüyle çaprazlanmış desenle, yuvalanmış deseni karşılaştırdıkları çalışmada, çaprazlamış desende G ve Phi katsayıları daha yüksek bulunmuştur. Birey ana etkisini varyans çaprazlanmış desen için daha yüksek, kalan etkiye ilişkin varyans değerini daha düşük bulmuşlardır. Araştırmacılar, sınıf içi uygulamalarda tümüyle çaprazlanmış modelin daha güvenilir sonuçlar verdiğini savunmuşlardır. Çalışmanın sonucunda, uygulanabildiği durumlarda, tümüyle çaprazlanmış modelin kullanılmasının daha uygun olacağını önermişlerdir.

Bir iş performansı ölçeği aracılığı ile elde edilen verilerin klasik test ve genellenebilirlik kuramına göre incelendiği çalışmada, test-tekrar test ve Cronbach alfa katsayısı klasik test kuramında güvenirlik amacıyla değerlendirilmiştir. Genellenebilirlik kuramına göre ise, G ve Phi katsayıları belirlenmiştir. Çalışma sonucunda, farklı güvenirlik katsayılarının benzer olmamasının klasik test kuramının sınırlı kalmasını gösterdiğini belirtilmiştir. İki kuramın benzer katsayıları vermesine karşın, genellenebilirlik kuramında tek bir analizle güvenirlik verilerine ulaşılmaktadır. Özellikle karar çalışması ile farklı senaryoların etkilerinin de analiz edilebilmesinin genellenebilirlik kuramının önemli avantajlarından olduğu belirtilmiştir (Yelboğa ve Tavşancıl, 2010).

İlköğretim öğrencilerinde matematik dersinde problem çözme becerilerinin araştırıldığı bir çalışmada, açık uçlu sorulara verilen yanıtlar, analitik ve bütünsel dereceli puanlama anahtarları ile farklı matematik öğretmenleri tarafından değerlendirilmiştir. Çalışmada, klasik test kuramı ve genellenebilirlik kuramına göre elde edilen veriler birbirine yakın bulunmuştur. Puanlayıcılar arası tutarlılık yüksek bulunmuş, Analitik ve bütünsel dereceli puanlama anahtarları arasındaki farkın analitik lehine görece daha yüksek olduğu belirtilmiştir Yıldıztekin, 2014).

Yıldırım Seheryeli'nin, yazılı anlatım becerisinde puanlama anahtarının puanlayıcılar arası güvenirliğinin incelendiği çalışmasında, klasik test kuramı ve genellenebilirlik kuramına göre değerlendirme yapılmıştır. Araştırmada, Genellenebilirlik Kuramının, maddelere ilişkin

hata tahminlerinde ve yetenek düzeyi belirlemede Klasik Test Kuramına göre daha detaylı bilgiler verdiği gösterilmiştir (Yıldırım Seheryeli, 2018).

Profesyonel futbolcu adayı 72 futbolcunun dört uzman tarafından değerlendirildiği bir çalışmada Futbol Yetilerine İlişkin Dereceleme Ölçeği kullanılmıştır. Araştırmada, ölçeğin güvenirliği incelenmiş ve Klasik test kuramı ve genellenebilirlik kuramına göre karşılaştırma yapılmıştır. Video görüntüleri ile değerlendirilen performansta, değerlendirmeler bir hafta ara ile tekrarlanmıştır. İki farklı zamandaki puanlamalar arası tutarlılığa Pearson katsayısı, maddelerin iç tutarlılık için ise Cronbach alfa katsayısı ile bakılmıştır. Genellenebilirlik kuramında G ve Phi katsayıları araştırılmıştır. Çalışmada, birden fazla hata kaynağının olduğu durumlarda genellenebilirlik kuramının klasik test kuramına alternatif olduğu belirtilmiştir (Deliceoğlu, 2009).

Spor performansının değerlendirildiği diğer bir çalışmada Öztürk (2011), voleybol becerisinde video görüntülerinin değerlendirilmesinde güvenirlik katsayıları belirlenmiştir. Çalışmada, klasik test kuramına göre iç tutarlılığın yeterli olduğu ancak, farklı zamandaki puanlamalar arasındaki güvenirliğin yeterli düzeyde olmadığı belirtilmiştir. Araştırma sonunda voleybol beceri performansının değerlendirilmesinde gözlem formu yerine dereceli puanlama anahtarının kullanılması önerilmiştir (Öztürk, 2011).

Performans değerlendirmesine dayalı uygulamaların sıkça yer verildiği sağlık eğitimi gibi alanlarda, çok değişkenli hata kaynaklarının aynı anda ele alınarak yapılan ölçümlerin ayrıntılı güvenilirlik analizine yönelik yapılan çalışmalar ülkemizde sınırlı sayıdadır. Pratik performansın değerlendirilmesinde, kullanılan ölçme araçları için yapılacak kapsamlı güvenilirlik analizlerinin ve çok değişkenli hata kaynaklarını ele alan çalışmaların önemli olduğu belirtilmektedir (Uzun vd., 2018; Yılmaz ve Başbaşa, 2015; Yılmaz ve Gelbal, 2011).

2.2. Sağlık alanında yapılan çalışmalar

Eğitimde performans değerlendirmeye yönelik güvenirlik çalışmalarında GK çalışmaları yapılmakla birlikte, sağlık alanında yapılan güvenirlik çalışmalarında yüksek oranda güvenirlik ve geçerlik beklenir. Gösterilen performans ve kullanılan ölçüm araçlarının

hata kaynaklarının iyi değerlendirilmesi ve ortaya konması gereklidir (Kurtz, Silveman and Draper, 1998).

Fizyoterapi ve rehabilitasyon alanında genellenebilirlik kuramına ilişkin çalışmaların, klinik değerlendirme ve uygulamalar ile ilgili olduğu görülmektedir. Bu çalışmalarda temel hedef olarak, klinikte hasta değerlendirilmesinde kullanılan yöntemlerin ve etki eden faktörlerin incelenmesine yöneliktir.

Gatti ve diğerlerinin çalışmasında (2020), araştırmacılar, fizyoterapi ve rehabilitasyon alanında kullanılan değerlendirme yöntemlerinden olan kas kuvveti ölçümünü çalışmışlardır. Çalışmada, gerçek olmayan diz ekstansiyon kas kuvveti (sanal) verileri üzerinde çalışılmıştır. Çalışmada, birey, ölçümün yapıldığı gün ve test sayısı yüzey olarak alınmıştır. Analiz sonucunda en büyük varyans yüzdesi bireye ait bulunmuştur (%93). Ancak verilerin gerçek hasta üzerinde olmaması, diğer yüzey etkilerinin hiç olmamasına (%0) neden olduğu açıktır. İleride yapılacak çalışmalara model olması amaçlanan bu araştırmada GT'nin, klinik araştırmacılar için hasta değerlendirirken oluşabilecek ölçümleriyle ilişkili hatanın ayrıntılı yorumlanmasını ve anlaşılmasını sağlayan bir araç olarak söz edilmektedir (Gatti vd., 2020)

Diğer bir çalışmada (görüş makalesi) (Preus, 2013) fizyoterapide klinik değerlendirme ve ölçüm protokollerinin oluşturulmasında, klasik test teoremine göre genellenebilirlik kuramının (GK) iki ana avantajından söz edilmektedir. Bunlardan ilki, GK'nın aynı anda birden fazla ölçüm hata kaynağını dikkate almasıdır. İkincisi ise, değerlendiricinin ilk verileri genelleştirilmesine ve farklı kombinasyonlarda yeniden ölçümlerle (karar çalışmaları) izin vermesidir. Özellikle hastalar üzerinde yapılacak ölçümlerde, kullanılacak ölçüm yöntemlerinin analizi sonucunda, klinik ölçüm hataları azaltılabilir ve daha az hasta üzerinde çalışmalarla daha gerçekçi sonuçlara ulaşılabilir.

Diş hekimliği öğrencilerinde iletişim becerilerinin değerlendirildiği çalışmada, en yüksek varyans yüzdesinin görev bileşeninde olduğu belirlenmiştir (Uzun ve diğerleri, 2018). Araştırmacılar, sağlık alanındaki iletişim becerilerine yönelik görevlerin, farklı zorluk düzeylerine sahip olabileceğini belirtmişlerdir. Geliştirilmesi hedeflenen performansın düzeyi hakkında bilgi vermeyi daha kolay hale getirecek değerlendirme çizelgelerinin yararlı olabileceğini savunmuşlardır.

Sağlık alanında verilen eğitimlerde, eğitimin her aşamasında (teorik, pratik ve klinik uygulamalar) mesleki yeterliliklerin, asgari seviyede karşılanıp karşılanmadığının değerlendirilmesi halen önemli bir sorun olarak devam etmektedir. Hemşirelik alanında yeterliliklerin ölçme ve değerlendirilmesinde belirtilen güvenirlik çalışmaları içerisinde Genellenebilirlik kuramı çalışmalarının oldukça az olduğu belirtilmektedir. Genellikle kullanılan güvenirlik çalışmalarının puanlayıcılar arası güvenirlik çalışmaları ile sınırlı olduğu ifade edilmektedir (Adamson vd., 2012; O'Brien vd., 2019).

Özellikle sağlık alanında GK son 10 yılda güvenirliğin araştırıldığı çalışmalarda kullanılmaya başlandığı görülmektedir. Buna rağmen eğitimde, özellikle performans değerlendirmesi ve simülasyon çalışmalarının değerlendirilmesinde halen yaygın kullanımı bulunmamaktadır. Hemşirelik alanında, simülasyon kullanılarak hemşirelik yeterliliğinin ölçülmesinde kullanımına dair çalışmalar son yıllarda yapılabilmiştir (Prion vd., 2016). Bir başka çalışmada, simülasyon çalışmalarının değerlendirilmesinde, yöntem ve araçların psikometrik gelişimine dair daha fazla ve ayrıntılı çalışmaların yapılması gerektiği belirtilmiştir (Mariani ve Doolen, 2016).

O'Brien ve arkadaşlarının (2019) yaptıkları, simülasyon çalışması değerlendirilmesinde GT ve Karar analizi sonucunda, puanlayıcıya/değerlendiriciye ait varyans yüzdesinin azaltılmasında puanlayıcı/değerlendirici eğitim yöntemlerine odaklanılması gerektiğini belirtmişlerdir. Kreiter ve Zaid (2020), sağlık alanında Genellenebilirlik kuramının daha fazla gelişmesi ve yenilikler kazandırılabilmesi için birlikte çalışmaların gerektiğini belirtmişlerdir.

Sağlıkta, özellikle tıp eğitiminde yapılandırılmış objektif klinik değerlendirmelerde en büyük varyans kaynağı olarak yeteneklerdeki bireysel farklılıklardan dolayı, öğrenciler gösterilmektedir. Puanlayıcılara veya diğer yönlerden kaynaklanabileceği düşünülen büyük miktarda varyans, öğrenci hakkındaki kararları etkilememesi gerektiğinden, istenmeyen bir durumdur. Sağlık eğitimiyle ilgili GK çalışmalarında en büyük varyans kaynağının test edilen konular olup olmadığının iyi değerlendirilmesi gerektiği belirtilmektedir. Bu alandaki çalışmalar, yalnızca yapılacak araştırmanın güvenirliğini için değil, aynı zamanda diğer araştırmacı ve eğitimcilere yarar sağlamak için de olacaktır (Monteiro vd., 2019).

Sağlık eğitiminde, öğrenci değerlendirme konusunda, klinik değerlendirme piramidinin üçüncü basamağı, performans değerlendirme (bilgi ve yetenek, gösterilecek tutum) kritik bir öneme sahiptir. Performans ağırlıklı eğitimlerde ilk etapta öğrenci veya öğretim elemanının hasta rolü ile yapılan ilk uygulamalar tercih edilir. Bu aşamaya yönelik değerlendirilmenin GK kuramına göre incelendiği çalışmalara rastlanamamıştır. Hasta üzerinde istasyon değerlendirmelerine yönelik kapsamlı çalışmalar yeterli sayıda değildir (Yılmaz ve Gelbal, 2011; Yılmaz ve Başbaşa, 2015; Yılmaz ve Tavşancıl, 2014; Uzun vd., 2018).

Fizyoterapi ve rehabilitasyon alanında ülkemizde yapılan bir çalışmada, son sınıf pratik sınavında analitik puanlama anahtarının puanlamaya etkisi incelenmiştir. Klasik test kuramına göre yapılan araştırmada, dereceli puanlama anahtarı ile yapılacak değerlendirmenin daha güvenilir sonuçlar vereceği sonucuna varılmıştır (Doğan ve Yosmaoğlu, 2015). Kas iskelet sistemi problemlerinde öğrenci performansı değerlendirilmiş ve her bir değerlendirmenin uzun sürmesi limitasyon olarak belirtilmiştir. Güney Afrika'da yapılan bir diğer çalışmada ise fizyoterapide pratik performans değerlendirmesi için eklem hareket açıklığının ölçülmesinde gonyometrik ölçüm performansı değerlendirilmiştir. Objektif yapılandırılmış pratik değerlendirme sonucunda, genel olarak puanlayıcılar arası güvenirliğin iyi-yüksek olarak gösterilmiştir. Ancak çalışmada, teknikte ve özellikle temel bilgi puanlamalarında, puanlayıcılar arasında istatistiksel farklar görülmektedir. Klasik test kuramına göre yapılan analizde, fizyoterapi eğitimcilerinin çok özel ve ayrıntılı değerlendirme kriterleri belirlemelerinin gerektiği belirtilmiştir (Olivier vd., 2013).

ÜÇÜNCÜ BÖLÜM

YÖNTEM

Bu bölümde araştırmanın modeli, verilerinin toplandığı grup, veri toplama araçları ve elde edilen verilerin analizinde kullanılan yöntemler belirtilmiştir.

3.1. Araştırma Modeli

Bu çalışmada, fizyoterapi ve rehabilitasyon öğrencilerinde eklem hareket sınırının gonyometrik ölçüm becerilerinin pratik değerlendirilmesi ile elde edilen ölçümlerin güvenirliğinin klasik test ve genellenebilirlik kuramlarına göre araştırılması amaçlanmıştır. Değerlendirmede oluşturulan, "dereceleme ölçeği"nin puanlayıcılar arasında uyumunun değerlendirilmesi hedeflenmiştir. Bunun yanında, yaygın olarak kullanılmakta olan "genel izlenim" ile performans değerlendirme yöntemi ile olan benzerlik ve farklılıkların incelenmesi amaçlanmıştır.

Temel araştırmalar, bilgi, kuram üretmeye yönelik çalışmalar olarak tanımlanmaktadır (Büyüköztürk vd., 2019). Yöntemsel analiz içeren araştırmalar bu kapsamda değerlendirilmektedir. Ayrıca, kuram ve hipotez üretmeye yönelik çalışmalar da bu içerikte kabul edilmektedir. Bu araştırmada puanlayıcılar, öğrencilerde dereceleme ölçeği ve genel izlenim ile öğrencilerde değerlendirme yapmışlar, her iki yöntem karşılaştırılmıştır. Çalışmada kullanılan dereceleme ölçeği iki farklı kurama göre incelenmiştir. Dolayısıyla, bu çalışma temel araştırma niteliğindedir.

3.2. Evren ve Örneklem

Araştırmanın evrenini, eklem hareket açıklığının universal gonyometre ile ölçümü eğitimini teorik ve pratik olarak almış fizyoterapi ve rehabilitasyon lisans eğitimi öğrencileri oluşturmaktadır. Öğrencinin pratik eğitimi almış olmalarına bakılmış, ders sonucunda başarılı olup olmadıkları kriter olarak alınmamıştır. Araştırmanın örnekleminde seçkisiz olmayan örnekleme yöntemlerinden uygun örnekleme yöntemi kullanılmıştır. Uygun örnekleme yönteminde zaman, para ile işgücü kaybının önlenmesi hedeflenir (Büyüköztürk vd., 2019). Örneklem amacıyla, Hasan Kalyoncu Üniversitesi Sağlık Bilimleri Fakültesi Fizyoterapi ve Rehabilitasyon Bölümü ve Sağlık Bilimleri Üniversitesi Gülhane Sağlık Bilimleri Fakültesi Fizyoterapi ve Rehabilitasyon Bölümü'nden çalışmaya gönüllü olarak katılmayı kabul eden

öğrenciler araştırmaya dâhil edilmiştir. Çalışmaya, Hasan Kalyoncu Üniversitesi'nden 33 ve Sağlık Bilimleri Üniversitesi'nden 20 olmak üzere toplam, 53 öğrenci araştırmaya katılmıştır. Çalışmaya kabul edilen puanlayıcılar performansın değerlendirildiği dersi vermekte olan öğretim elemanlarından oluşturulmuştur. Hasan Kalyoncu Üniversitesi'nden iki, Sağlık Bilimleri Üniversitesi'nden bir, Hacettepe Üniversitesi'nden bir, Doğu Akdeniz Üniversitesi'nden bir ve Haliç Üniversitesi'nden bir puanlayıcı çalışmaya gönüllü olarak katılmıştır. Puanlayıcıların bu performansın yer aldığı dersteki deneyimleri, 5-15 yıl arasında değişmekteydi.

3.3. Veri Toplama Araçları

3.3.1. Eklem Hareket Sınırının Değerlendirilmesinde Dereceleme Ölçeği

Ölçekte, el bileği eklemi ekstansiyon hareketi sınırının universal gonyometre ile ölçülmesi hedeflenmiştir. Eklem hareket sınırının ölçülmesi amacıyla dereceleme ölçeği hazırlanmıştır. Ölçek için öncelikle, temel performans kriterleri belirlenmiştir (Norkin ve White, 2016; Olivier vd, 2013; Myezwa, 2013). Toplamda 10 maddeden oluşan gonyometrik ölçüm becerisine yönelik görev, temel kaynak ve temel ölçme değerlendirme ders içeriğine göre belirlenmiştir (Carroll University, 2013; University of Minnesota, 2019). Ölçeğin madde sayısı ilgili uzman ekip tarafından 8 maddede toplanmıştır.

Geliştirilen dereceleme ölçeğinde her görev için, 0: "Yok", 1: "Yetersiz", 2: "Orta", 3: "Yeterli" puanlama dereceleri kullanıldı.

"Yok" derecesi: Hiç veya yanlış uygulama veya cevaplamayı,

"Yetersiz" derecesi: Yetersiz uygulama veya cevaplamayı,

"Orta" derecesi: Orta derecede uygulama veya cevaplamayı,

"Yeterli" derecesi: İstenen veya beklenen düzeyde uygulama veya cevaplamayı ifade etmektedir.

Ölçekteki kullanılan görevler aşağıdaki gibi oluşturulmuştur:

Seviyeler:	Yok	Yetersiz	Orta	Yeterli
Görevler:				
1. Hastanın pozisyonlanması	☐	☐	☐	☐
2. Gonyometrenin türünün bilinmesi ve gonyometrenin sabit ve hareketli kolların yeri	☐	☐	☐	☐
3. Pivotun yeri/palpasyonu, gonyometrenin yerleştirilmesi	☐	☐	☐	☐
4. Hastaya yapılacak işlem hakkında açıklamada bulunulması	☐	☐	☐	☐
5. Ölçücünün hareketi göstermesi ve kendini pozisyonlaması	☐	☐	☐	☐
6. Ölçüm için destek ve stabilizasyon	☐	☐	☐	☐
7. Gonyometrik ölçüm işlemi ve aşamaları	☐	☐	☐	☐
8. Eklem hareket sınırı değerleri ve gonyometrenin okunması bilgisi	☐	☐	☐	☐

Çalışmada aynı puanlayıcılar tarafından kullanılan diğer bir puanlama yöntemi ise puanlamanın bir bütün olarak yapıldığı "genel izlenim" yöntemi ile puanlama olmuştur. Puanlayıcılardan, deneyimlerine dayanarak yukarıdaki maddelerin sergilendiği performansta öğrenciye genel izlenim sonucunda tek bir puan vermesi (100 üzerinden) istenmiştir. Bütünsel bir puanlamanın yapıldığı genel izlenimde, puanlayıcıdan performansın tamamını izleyerek genel olarak tek bir puan vermesi istenmiştir.

3.3.2. Kapsam Geçerliği Çalışması

Oluşturulan dereceleme ölçeğinin görevleri (eklem hareket sınırının gonyometre ile ölçülmesi), temel eğitim kaynak kitapları ve farklı eğitim kurumlarının müfredatlarından alınan standart görevler taranmış (Norkin ve White, 2016; Olivier vd, 2013; Myezwa, 2013; Carroll University, 2013; University of Minnesota, 2019), her görev için bu alanda eğitim veren deneyimli 5 öğretim üye ve görevlisinin görüşleri ile dereceleme ölçeğine 8 maddeden

oluşan son şekli verilmiştir. Ölçek, gonyometrik ölçümün yapılabileceği tüm eklemlerde de kullanılabilecek şekilde standardize edilmiştir.

3.3.3. Güvenirlik Çalışmaları

Dereceleme ölçeğinin test-tekrar test güvenirliği için Puanlayıcı 1'den tüm öğrencilerden bu değerlendirmeyi, yapılacak ikinci değerlendirmedeki puanlamalar arası zamanın puanlama arasındaki etkisini en aza indirmek amacıyla, 1 hafta sonra tekrar yapması istenmiştir. Benzer işlem genel izlenime göre puanlama amacıyla Puanlayıcı 2'den istenmiştir. Test-tekrar test güvenirliği Sınıf-içi korelasyon katsayısı (Intraclass Correlation Coefficient, ICC) ile incelenmiştir. ICC 2,1 modeli ile, ICC katsayısı, zayıf (<0.5), orta (0.5-0.75), iyi (0.75-.90) ve mükemmel (>0.90) olarak derecelendirilmiştir (Irwin, 2019). Dereceleme ölçeğinin 6 puanlayıcı arasındaki güvenirliğe Sınıf-içi korelasyon katsayısı ve Pearson korelasyon analizi ile bakılmıştır.

Dereceleme ölçeğinin iç tutarlılığının belirlenmesinde Cronbach Alfa katsayısı kullanılmıştır. Alfa katsayısı için 0,70'i kabul edilebilir, 0,80 üzeri tercih edilebilir, 0,90 ve üzerinin ise mükemmel olarak belirtilmektedir (Cotina, 1993; Taber, 2018).

Altı puanlayıcının 8 göreve (maddeye) ilişkin verdikleri puanlar için genellenebilirlik kuramında, "birey x puanlayıcı x madde, (b x p x m)" çapraz deseni kullanılarak değişkenlik kaynakları incelenerek güvenirlik analizleri yapılmıştır.

3.3.4. Veri Toplama Araçlarının Uygulanması

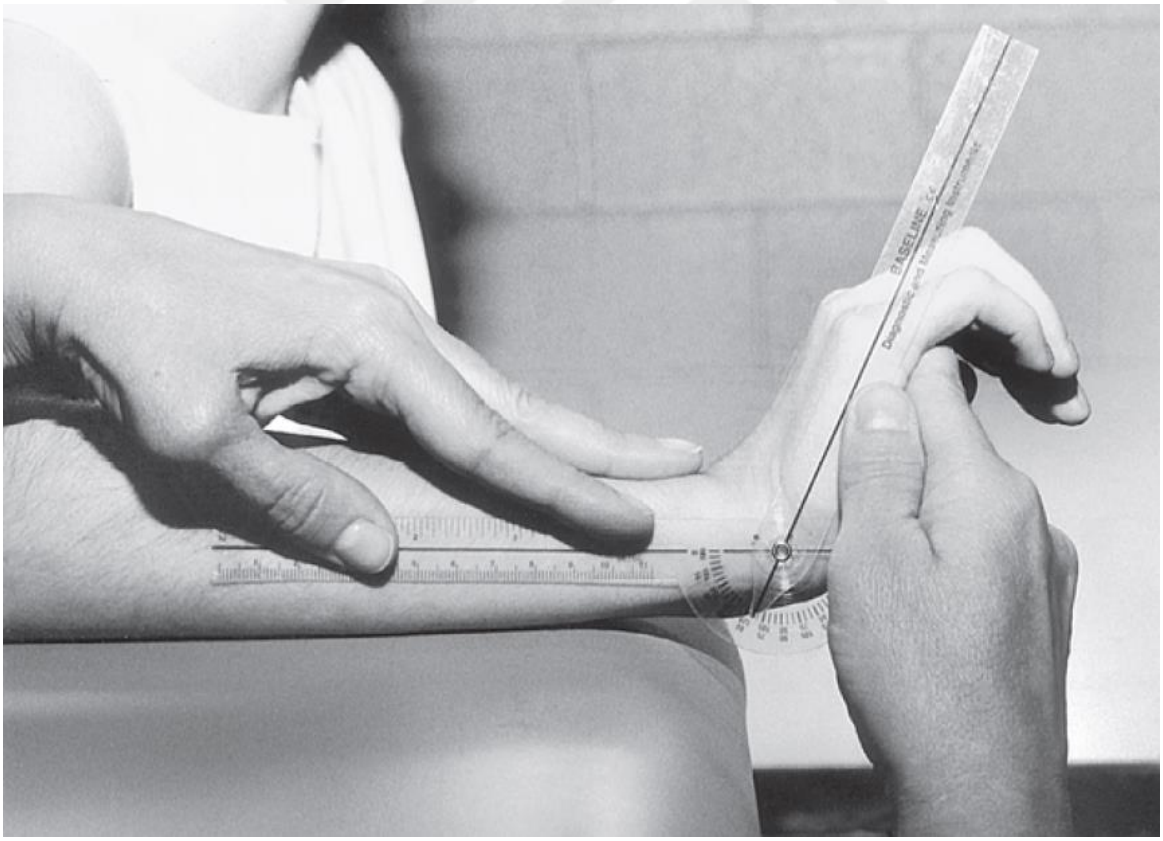
Bu araştırmada, Hasan Kalyoncu Üniversitesi Sağlık Bilimleri Fakültesi ve Sağlık Bilimleri Üniversitesi Gülhane Sağlık Bilimleri Fakültesi Fizyoterapi ve Rehabilitasyon bölümleri öğrencilerinin, el bileği eklem hareket açıklığının universal gonyometre ile ölçülmesi pratik performansları görüntülerinden elde edilen veriler kullanılmıştır.

El bileği hareket sınırını göstermesi ve aktif ve pasif hareket sınırının ne kadar olduğunu (60-70 derece) belirtmesi istenmektedir (Şekil 3.1) (8. görev, m8). Ölçüm için kullanılacak olan ölçüm aracının (gonyometre) türünü (universal gonyometre) (Şekil 3.2) ifade etmesi beklenmektedir (Görev 2). El bileği eklem hareket açıklığının universal

gonyometre ile ölçülmesinde, öğrenciden ölçüm yapılacak kişiden (rol öğrenci) el bileğini bir masa üzerine yerleştirmesi ve uygun pozisyonu vermesi beklenmektedir (1. görev, m1). Gonyometrenin sabit ve hareketli kolların yerleştirilmesi ve nerede olduklarının gösterilmesi (2. görev, m2), palpe edilerek pivotun gösterilmesi ve gonyometrenin yerleştirilmesi (3. görev, m3), ölçüm yapılacak kişiye yapılacak işlem hakkında açıklamada bulunulması (4. görev, m4), hareketi göstermesi ve öğrencinin kendi pozisyonunu alması (5. görev, m5), gerekli destek ve stabilizasyonun sağlanması (6. görev, m6), eklem hareket sınırının ölçülmesi (7. görev, m7) ve en sonunda gonyometreden elde edilen değerin okunması (8. görev, m8) değerlendirilmektedir.

Şekil 2: El bileği ekstansiyon hareketi

Şekil 3: Universal gonyometre



Şekil 4: El bileği ekstansiyon hareketinin universal gonyometre ile ölçümü

Eklem hareket sınırının universal gonyometre ile ölçülmesinde temel ölçütler:

- Hareketin adı: El bileği ekstansiyonu.
- Test (ölçüm) pozisyonu: Oturur pozisyonda. Omuz eklemi 90 derece abduksiyonda, dirsek eklemi 90 derece fleksiyonda, avuç içi yere bakacak şekilde önkol nötralde. Önkol destekli olacak şekilde, el, el bileğinden itibaren boşta. El bileği ne ulnar, ne de radial deviasyonda. Parmaklar gevşek.
- Destek (stabilizasyon): Dirsek ve önkol hareketlerini önlemek amacıyla radius ve ulna destekli.
- Pivot: El bileği lateral kenarı, triquetrum üzerinde.
- Proksimal kol (sabit kol): Ulnanın lateral kenarı. Olekranon ve ulnanın styloid çıkıntısı referans alınır.
- Distal kol (hareketli kol): Beşinci metakarpalin lateral orta hattı.

Araştırmaya toplam 53 öğrenci dahil edilmiştir (b1-53). Puanlayıcılar farklı üniversitelerde eğitim veren öğretim elemanlarından oluşturulmuştur (Puanlayıcı 1-6). Öğrenci pratik performansları kameraya sesli kayıt olarak alınmıştır.

Tüm puanlayıcılardan öğrenci performansını önce genel izlenim ile değerlendirmesi istenmiştir. Genel izlenim puanlama yapısı Tablo 3'te gösterilmiştir.

Değerlendirmeden en az 1 hafta sonra, yine tüm puanlayıcılardan bu kez, oluşturulan dereceleme ölçeği puanlama anahtarı ile öğrenci performanslarını tekrar değerlendirmeleri istendi. Öğrencilerin (b1-53) performanslarına, puanlayıcılar tarafından (Puanlayıcı 1-6). Dereceleme ölçeği puanlama anahtarındaki görevlere (m1-8), öğrenci performansının seviyesine göre puan vermeleri istendi (x: Yok (0), Yetersiz (1), Orta (2), Yeterli (3)). Araştırmada kullanılan birey (öğrenci) (b), puanlayıcı (p) ve madde (görev) (m) deseni aşağıdaki Tablo 4'te verilmiştir.

Tablo 3: Genel izlenim ile performans değerlendirme veri yapısı

	Öğrenci	Genel izlenim
Puanlayıcı 1	b1	.
(Nb=53)	b2	.
	…	.
	b53	.
Puanlayıcı 2	b1	.
(Nb=53)	b2	.
	…	.
	b53	.
Puanlayıcı 3	b1	.
(Nb=53)	b2	.
	…	.
	b53	.
Puanlayıcı 4	b1	.
(Nb=53)	b2	.
	…	.
	b53	.
Puanlayıcı 5	b1	.
(Nb=53)	b2	.
	…	.
	b53	.
Puanlayıcı 6	b1	.
(Nb=53)	b2	.
	…	.
	b53	.

Tablo 4: İki yüzeyli çaprazlanmış b x p x m deseni veri yapısı

		Madde				
	Öğrenci	**m1**	**m2**	**…....**	**m7**	**m8**
Puanlayıcı 1	b1	·	·	….....	·	·
(Nb=53)	b2	·	·	….....	·	·
	…	·	·	….....	·	·
	b53	·	·	….....	·	·
Puanlayıcı 2	b1	·	·	….....	·	·
(Nb=53)	b2	·	·	….....	·	·
	…	·	·	….....	·	·
	b53	·	·	….....	·	·
Puanlayıcı 3	b1	·	·	….....	·	·
(Nb=53)	b2	·	·	….....	·	·
	…	·	·	….....	·	·
	b53	·	·	….....	·	·
Puanlayıcı 4	b1	·	·	….....	·	·
(Nb=53)	b2	·	·	….....	·	·
	…	·	·	….....	·	·
	b53	·	·	….....	·	·
Puanlayıcı 5	b1	·	·	….....	·	·
(Nb=53)	b2	·	·	….....	·	·
	…	·	·	….....	·	·
	b53	·	·	….....	·	·
Puanlayıcı 6	b1	·	·	….....	·	·
(Nb=53)	b2	·	·	….....	·	·
	…	·	·	….....	·	·
	b53	·	·	….....	·	·

Bu araştırmanın etik onayı Hasan Kalyoncu Üniversitesi Sağlık Bilimleri Fakültesi Girişimsel Olmayan Araştırmalar Etik Kurulu tarafından 28.05.2020 tarih ve 2020/034 sayılı kararı ile alınmıştır.

3.4. Verilerin Analizi ve Yorumlanması

Veriler, aritmetik ortalama ve standart sapma olarak ifade edilmiştir. Ayrıca minimum ve maksimum değerleri ve çarpıklık ve basıklık değerleri verilmiştir. Dereceleme ölçeğindeki her madde için elde edilen veriler medyan ve minimum-maksimum olarak verilmiştir.

Dereceleme ölçeğinde elde edilen veriler, çalışmada karşılaşılan zorunluluk nedeniyle, küçük bir gruptan (n=53) elde edilmiştir. Bu sınırlılık nedeniyle ve madde sayısının az olması nedeniyle (m=8) faktör analizi yapılmıştır. Ayrıca, hazırlanan dereceleme ölçeğinde uzmanlar görüş olarak maddeler arasında ilişki olduğunu bildirmişlerdir. Bu nedenle her puanlayıcının dereceleme ölçeği faktör analizi ile incelenmiştir.

Verilerin analizinde 3 farklı değerlendirme yapılmıştır: 1. Dereceleme ölçeği ve Genel izlenim puanlama değerlerinin karşılaştırılması. 2. Dereceleme ölçeğinin güvenirliğinin değerlendirilmesi (Test-tekrar test, puanlayıcılar arası, iç tutarlılık) ve 3. Dereceleme ölçeğinin genellenebilirlik kuramı ile analizi.

1. Dereceleme ölçeği ve Genel izlenim puanlama değerlerinin karşılaştırılması: Tüm puanlayıcılardan elde edilen genel izlenim değerleri (100 üzerinden) ve dereceleme ölçeği değerleri (10 görev, her biri 0-3 puan, toplam 30 puan. Öğrenci performans notu = Öğrencinin aldığı not x 100 / 30) her bir puanlayıcı için t testi ile karşılaştırılmış ve Pearson korelasyon analizi ile ilişkisi incelenmiştir.

2. Dereceleme ölçeği değerlendirmesinin güvenirliğinin değerlendirilmesi (Test-tekrar test): Puanlayıcı 1'den analitik dereceli puanlamayı tüm öğrenciler için (Nb=53) bir hafta arayla tekrar değerlendirmesi istenmiştir. İki değerlendirme arasındaki test-tekrar test güvenirliği ICC ile değerlendirilmiştir. Dereceleme ölçeğinin, puanlayıcılar (6 puanlayıcı) arası değerlendirilmesinde puanlayıcılar arası güvenirlik için benzer şekilde, Sınıf içi korelasyon katsayısı ICC kullanılmıştır. Dereceleme ölçeğinin iç tutarlılık geçerliğinin değerlendirilmesi Cronbach Alfa ile yapılmıştır.

Yukarıdaki 3 maddedeki istatistiksel analiz için IBM SPSS Statistics, Version 25 programı kullanılmıştır.

3. Analitik dereceli puanlama anahtarının genellenebilirlik kuramı analizi: Verilerin analizinde b x p x m desenine göre genellenebilirlik kuramı analizi yapılmıştır. Araştırmada, 6 puanlayıcı tarafından, tüm öğrencileri pratik performanslarının, bu araştırmada oluşturulan analitik dereceli puanlama ölçeğine göre değerlendirmelerinin yapılması istenmiştir. Tamamı çaprazlanmış b x p x m desenine göre genellenebilirlik çalışması yapılmıştır. Tamamı çaprazlanmış iki yüzeyli b x p x m deseninin, genellenebilirlik kuramına göre değerlendirilmesi amacıyla EduG programının 6.1 e versiyonu kullanılmıştır (EduG - 6.1 e). Genellenebilirlik kuramı sonrasında Karar çalışması için de aynı program kullanılmıştır.

DÖRDÜNCÜ BÖLÜM

BULGULAR VE TARTIŞMA

Bu bölümde, dereceleme ölçeği ve genel izlenim değerlendirme yöntemleriyle 53 öğrencide (birey) yapılan ve 6 değerlendirici (puanlayıcı) ile değerlendirilen bulgularla ilgili veriler ve yorumları yer almaktadır. İlk bölümde her iki puanlama ile bireylerden elde edilen verilerin betimsel istatistiksel sonuçlarına yer verilmiştir.

4.1 Betimsel İstatistikler

Çalışmaya katılan iki farklı puanlama yöntemi ile verdikleri notlara ilişkin veriler Tablo 4, Tablo 5 ve Tablo 6'da gösterilmiştir.

Tablo 5. Dereceleme Ölçeğine göre yapılan puanlamaya ait betimsel istatistikler

	Ortalama	Standart sapma	Minimum	Maksimum	Çarpıklık	Basıklık
1. Puanlayıcı	68,64	20,52	8,3	95,8	-0,965	0,780
2. Puanlayıcı	80,58	17,30	29,2	100,0	-1,019	0,520
3. Puanlayıcı	55,43	28,77	4,2	100,0	0,282	-1,372
4. Puanlayıcı	66,67	24,76	20,8	100,0	-0,025	-1,422
5. Puanlayıcı	82,22	16,36	41,7	100,0	-0,922	-0,232
6. Puanlayıcı	65,49	21,26	16,7	95,8	-0,320	-0,972

Puanlayıcılarda çarpıklık katsayıları dereceleme ölçeği için (-1,019 ile 0,282 arasında) ve genel izlenim için (-0,709 ile -0,121 arasında) incelendiğinde değerlerden yalnızca biri hariç (-1,019) -1 ile +1 arasında olduğu görülmektedir. Basıklık katsayıları incelendiğinde ise,

dereceleme ölçeği için (-1,422 ile 0,520 arasında) ve genel izlenim için (-1,327 ile -0,239 arasında) değerlerin -1,5 ile +1 arasında olduğu görülmektedir. Bu sonuçlara göre, verilerin dağılımının çarpık olmadığı, basıklığının ise kabul edilebilir düzeyde kaldığı söylenebilir.

Tablo 6. Dereceleme ölçeğine göre maddelere ait istatistikler (medyan ve minimum-maksimum)

	Puanlayıcı					
	1.	2.	3.	4.	5.	6.
Madde 1	2 (0-3)	3 (0-3)	2 (0-3)	3 (0-3)	3 (0-3)	3 (0-3)
Madde 2	2 (0-3)	2 (0-3)	1 (0-3)	3 (0-3)	2 (0-3)	2 (0-3)
Madde 3	3 (0-3)	3 (0-3)	1 (0-3)	2 (0-3)	3 (0-3)	2 (0-3)
Madde 4	3 (0-3)	3 (1-3)	2 (0-3)	1 (0-3)	3 (1-3)	3 (0-3)
Madde 5	3 (0-3)	3 (1-3)	1 (0-3)	2 (0-3)	2 (1-3)	2 (0-3)
Madde 6	2 (0-3)	3 (0-3)	1 (0-3)	2 (0-3)	3 (0-3)	2 (0-3)
Madde 7	2 (0-3)	2 (0-3)	1 (0-3)	2 (1-3)	3 (1-3)	2 (0-3)
Madde 8	1 (0-3)	3 (0-3)	1 (1-3)	2 (0-3)	3 (0-3)	1 (0-3)

Tablo 7. Genel izlenim puanlamasına göre yapılan puanlamaya ait istatistikler

	Ortalama	Standart sapma	Minimum	Maksimum	Çarpıklık	Basıklık
1. Puanlayıcı	67,70	21,59	15,0	98,0	-0,709	-0,239
2. Puanlayıcı	71,70	20,00	25,0	100,0	-0,464	-0,468
3. Puanlayıcı	62,45	21,32	15,0	100,0	-0,307	-0,370
4. Puanlayıcı	72,87	17,46	35,0	100,0	-0,240	-0,528
5. Puanlayıcı	80,94	13,38	50,0	100,0	-0,241	-0,520
6. Puanlayıcı	72,69	16,86	40,0	97,5	-0,121	-1,327

Tablo 7'de dereceleme ölçeğine göre en yüksek ortalama 5. Puanlayıcı'ya ait bulunmuştur (82,22). En düşük ortalama ise 3. Puanlayıcı'ya aittir (55,43). Genel izlenim puanlaması sonuçlarına göre, benzer şekilde, en yüksek ortalama 5. Puanlayıcı'ya ait (80,94), en düşük ortalama ise 3. Puanlayıcı'ya ait bulunmuştur (62,45). Dereceleme ölçeğinde puanlayıcılar arasındaki karşılaştırmada (bağımlı gruplarda t testi ile) 1.-4., 1.-6. ve 4.-6. puanlayıcı dışındaki diğer tüm değerlendirici eşleştirmelerinde fark bulunmuştur ($p<0,05$). Genel izlenim puanlamalarında, puanlayıcılar arasındaki karşılaştırmada ise (bağımlı gruplarda t testi ile) 1.-2., 2.-4., 2.-6. ve 4.-6. puanlayıcı dışındaki diğer tüm değerlendirici eşleştirmelerinde fark bulunmuştur ($p<0,05$).

Çalışmada dereceleme ölçeğinin faktör analizi sonuçları her puanlayıcı için ayrı ayrı yapılmış ve elde edilen veriler Tablo 8'de verilmiştir.

Tablo 8. Puanlayıcıların dereceleme ölçeği faktör analiz sonuçları

	Puanlayıcılar					
Maddeler	1.	2.	3.	4.	5.	6.
A1	0,341	0,499	0,711	0,649	0,486	0,551
A2	0,677	0,751	0,907	0,671	0,592	0,792
A3	0,647	0,713	0,824	0,696	0,592	0,808
A4	0,752	0,318	0,867	0,749	0,602	0,821
A5	0,54	0,604	0,841	0,701	0,537	0,575
A6	0,589	0,617	0,951	0,919	0,574	0,752
A7	0,315	0,629	0,942	0,82	0,837	0,809
A8	0,573	0,671	0,872	0,775	0,723	0,314
K-M-O	0,702	0,744	0,874	0,852	0,756	0,857
Bartlett X2	130,585	143,966	463,462	256,529	159,264	193,428
Öz değer	3,234	3,591	6,252	4,935	3,707	4,318
Açık. Varyans	40,42	44,881	78,149	61,684	46,335	53,973
Toplam A.V.	32,795	37,659	75,269	56,578	39,265	48,568

KMO: Kaiser-Meyer-Olkin.

Tabloda da görüldüğü gibi tüm puanlayıcılar için açıklanan varyans tek faktör için, 40'ın üzerindedir. Tüm maddelerin yükleri 0,30'un üzerinde hesaplanmıştır.

Çalışmaya katılan puanlayıcıların ortalama dereceleme ölçeği ve ortalama genel izlenim puanlamalarına ilişkin betimsel istatistikler Tablo 9'da verilmiştir.

Tablo 9. Çalışmaya katılan puanlayıcıların Dereceleme Ölçeği ve Genel İzlenim puanlamalarına ait betimsel istatistikler

	Dereceleme ölçeği	Genel izlenim
Ortalama	69,84	71,39
Standart sapma	18,72	16,45
Minimum	22,92	30,00
Maksimum	97,22	95,00
Varyans	350,521	270,479
Çarpıklık	-0,369	-0,323
Basıklık	-0,635	-0,458

Puanlayıcılardan elde edilen ortalamalar karşılaştırıldığında (bağımlı gruplarda t testi ile) iki puanlama yöntemi arasında fark bulunmamıştır (t=1,928; p=0,059).

4.2. Dereceleme ölçeği ile genel izlenim puanlarının test-tekrar test güvenirliğinin incelenmesi

Çalışmada 2. Puanlayıcı dereceleme ölçeğini, 1. Puanlayıcı ise genel izlenim puanlamasını 10 gün sonra tekrar yapmışlardır. 2. Puanlayıcının test-tekrar test güvenirliği her madde ve toplam puan için Sınıf İçi Korelasyon Katsayısı ile bakılmıştır. 1. Puanlayıcının test-tekrar test güvenirliği ise Sınıf İçi Korelasyon Katsayısı ile bakılmıştır. Test-tekrar test güvenirlik değerleri Tablo 10'da verilmiştir.

Tablo 10. 1. Puanlayıcının Genel İzlenim ve 2. Puanlayıcının Dereceleme Ölçeği test-tekrar test betimsel istatistikleri

	Dereceleme Ölçeği (2. Puanlayıcı)		Genel İzlenim (1. Puanlayıcı)	
	Test	Tekrar test	Test	Tekrar test
Ortalama	63,96	64,59	67,70	67,17
Standart sapma	14,01	13,81	21,59	20,91
Minimum	23,3	23,3	15	15
Maksimum	80	80	98	95
Varyans	196,178	190,695	466,292	437,028
Çarpıklık	-0,939	-0,996	-0,709	-0,640
Basıklık	0,243	0,535	-0,239	-0,339

Her iki puanlama türüne ait sınıf içi korelasyon katsayısı değerleri ve dereceleme ölçeğine ait maddelerin sınıf içi korelasyon katsayısı değerleri Tablo 11 ve Tablo 12'de verilmiştir.

Tablo 8. Dereceleme ölçeğine ait maddelerin test-tekrar test Sınıf içi korelasyon katsayısı değerleri

		%95 Güven aralığı	
	ICC	Alt	Üst
Madde 1	0,988	0,979	0,993
Madde 2	1,000	1,000	1,000
Madde 3	0,988	0,979	0,993
Madde 4	0,840	0,722	0,907
Madde 5	0,909	0,843	0,948
Madde 6	0,941	0,897	0,966
Madde 7	0,987	0,977	0,992
Madde 8	0,982	0,969	0,990

ICC Sınıf İçi Korelasyon katsayısı (Intraclass Correlation Coefficient).

Tablo 9. Dereceleme Ölçeği ve Genel İzlenim puanlama türlerine ait test-tekrar test Sınıf İçi Korelasyon katsayısı değerleri

		%95 Güven aralığı	
	ICC	Alt	Üst
Dereceleme Ölçeği	0,986	0,975	0,992
Genel İzlenim	0,988	0,980	0,993

ICC: Sınıf İçi Korelasyon katsayısı (Intraclass Correlation Coefficient).

Dereceleme ölçeğinde, maddelerin Sınıf içi korelasyon katsayısı (ICC) değerleri 0,840-1,000 arasında değişmektedir. Madde 4 dışındaki tüm maddelerde ICC değerleri mükemmeldir (ICC>0,90). Madde 4 iyi bulunmuştur (ICC=0,840). Dereceleme Ölçeği ve Genel İzlenim puanlamalarının test-tekrar test güvenirliği mükemmel bulunmuştur (ICC>0,90). Her iki puanlama türünde de test-tekrar test verileri bağımlı gruplarda t testi ile karşılaştırıldığında, fark bulunmamıştır (Dereceleme Ölçeği için: t=-1,938, p=0,058 ve Genel İzlenim için: t=1,190, p=0,239).

Dereceleme Ölçeği ve Genel İzlenim puanlamaların test-tekrar test puanları arasındaki korelasyon incelendiğinde, Dereceleme Ölçeği için r=0,986 (p<0,001) ve Genel İzlenim için r= 0,989 (p<0,001) olarak bulunmuştur.

4.3. Dereceleme ölçeği ve genel izlenim değerlendirmelerinde puanlayıcılar arası güvenirliğinin incelenmesi

Puanlayıcıların dereceleme ölçeği maddeleri ve toplam puanları ile genel izlenim puanlamaları arasındaki güvenirlik değerlendirmeleri Tablo 14'te verilmiştir.

Tablo 10. Dereceleme ölçeği maddeleri ve toplam puanları ile genel izlenim puanlamalarının puanlayıcılar arası güvenirliği

		%95 Güven aralığı	
	ICC	Alt	Üst
Madde 1	0,831	0,749	0,893
Madde 2	0,919	0,880	0,949
Madde 3	0,900	0,852	0,937
Madde 4	0,772	0,661	0,855
Madde 5	0,766	0,652	0,851
Madde 6	0,811	0,720	0,880
Madde 7	0,825	0,741	0,889
Madde 8	0,814	0,724	0,882
Toplam (Dereceleme Ölçeği)	0,926	0,890	0,953
Genel İzlenim	0,943	0,915	0,964

ICC: Sınıf İçi Korelasyon katsayısı (Intraclass Correlation Coefficient).

Sınıf içi korelasyon katsayısı değerleri maddeler için 0,766-0,919 arasında değişmektedir. En yüksek değer Madde 2 ve en düşük değer Madde 5'te gözlenmiştir. Tüm maddelerde Sınıf içi korelasyon katsayısı değerleri 1., 3., 4., 5., 6., 7. ve 8. maddeler için iyi (ICC: 0,75-0,90), 2. madde için ise mükemmel bulunmuştur (ICC>0,90). Dereceleme ölçeği puanlamasının toplamı ve genel izlenim puanlamasının puanlayıcılar arası güvenirliği mükemmel bulunmuştur (ICC>0,90).

4.4. Dereceleme ölçeğinin iç tutarlılığının incelenmesi

Dereceleme ölçeğinin iç tutarlılığının değerlendirilmesinde, Cronbach Alfa katsayısının 0,813-0,958 arasında değiştiği görülmüştür (Tablo 14). Alfa katsayısı dört puanlayıcıda 0,80'in üzerinde (tercih edilebilir düzeyde) ve iki puanlayıcıda da 0,90'ın üzerinde (mükemmel) bulunmuştur.

Tablo 14. Dereceleme Ölçeğine göre yapılan puanlamanın Cronbach Alfa değerleri

	Cronbach Alfa
1. Puanlayıcı	0,813
2. Puanlayıcı	0,821
3. Puanlayıcı	0,958
4. Puanlayıcı	0,902
5. Puanlayıcı	0,830
6. Puanlayıcı	0,874

4.5. Dereceleme ölçeği ile genel izlenim puanlamasında puanlayıcılar arası ilişkinin incelenmesi.

Puanlayıcıların dereceleme ölçeği toplam puanları için puanlayıcılar arasındaki ikilli korelasyonlar Tablo 15'te, aynı analizin genel izlenim puanlamaları için hesaplanan değerleri Tablo 16'da verilmiştir.

Tablo 11. Dereceleme ölçeği puanlarının puanlayıcılar arası korelasyonları

	1. Puanlayıcı	2. Puanlayıcı	3. Puanlayıcı	4. Puanlayıcı	5. Puanlayıcı	6. Puanlayıcı
1. Puanlayıcı	-	0,748*	0,681*	0,688*	0,693*	0,807*
2. Puanlayıcı		-	0,771*	0,743*	0,892*	0,816*
3. Puanlayıcı			-	0,910*	0,833*	0,848*
4. Puanlayıcı				-	0,814*	0,879*
5. Puanlayıcı					-	0,821*
6. Puanlayıcı						-

* p<0,001. r: Pearson korelasyon katsayısı.

Tablo 12. Genel izlenim puanlamalarının puanlayıcılar arası korelasyonları

	1. Puanlayıcı	2. Puanlayıcı	3. Puanlayıcı	4. Puanlayıcı	5. Puanlayıcı	6. Puanlayıcı
1. Puanlayıcı	-	0,683*	0,709*	0,717*	0,795*	0,859*
2. Puanlayıcı		-	0,786*	0,841*	0,790*	0,672*
3. Puanlayıcı			-	0,802*	0,820*	0,648*
4. Puanlayıcı				-	0,777*	0,692*
5. Puanlayıcı					-	0,768*
6. Puanlayıcı						-

* p<0,001. r: Pearson korelasyon katsayısı.

Puanlayıcılar arasında dereceleme ölçeği puanlamasında, korelasyon katsayısı 0,681-0,910 arasında değişmektedir. Genel izlenim puanlamasında ise korelasyon katsayısı 0,648-0,859 arasında değişmektedir. Tüm verilerde p<0,001 olarak bulundu.

Her bir puanlayıcının dereceleme ölçeği ve genel izlenim değerlendirmeleri arasındaki ilişki Tablo 17'de verilmiştir.

Tablo 13. Her bir puanlayıcının dereceleme ölçeği ve genel izlenim değerlendirmeleri arasındaki korelasyonları

	r
1. Puanlayıcı	0,869*
2. Puanlayıcı	0,924*
3. Puanlayıcı	0,804*
4. Puanlayıcı	0,674*
5. Puanlayıcı	0,860*
6. Puanlayıcı	0,970*

* p<0,001. Pearson korelasyon analizi.

Puanlayıcılarda dereceleme ölçeği ve genel izlenim puanlamaları ilişkileri 0,970 ile 0,674 arasında değişmektedir. En yüksek ilişki 6. Puanlayıcıda ve en düşük ilişki ise 4. Puanlayıcıda belirlenmiştir. Tüm değerlerde p<0,001 olarak bulunmuştur.

4.6. Dereceleme ölçeği puanları için G çalışması

Toplam 53 öğrenci için (**b**, B1-B53), 6 puanlayıcıdan (**p**, Puanlayıcı 1-6) ve 8 maddeye (**m**, Madde1-8) ilişkin elde edilen puanlar Genellenebilirlik kuramına göre analiz edilmiştir. Çalışma "b x p x m" desenine göre gerçekleştirilmiştir. Değişkenlik kaynaklarına ait değerler Tablo 18'de verilmiştir.

Tablo 14. Dereceleme ölçeği ile elde edilen verilerin b x p x m desenine göre G çalışmasıyla varyans bileşenleri ve varyans yüzdeleri

Varyans kaynağı	Kareler toplamı (SS)	Serbestlik Derecesi	Kareler ortalaması (MS)	Kestirilen varyans	Toplam Varyans Yüzdesi
b	787,52123	52	15,14464	0,28009	29,9
p	193,56447	5	38,71289	0,08150	8,7
m	38,64623	7	5,52089	0,00499	0,5
bp	291,14387	260	1,11978	0,10040	10,7
bm	326,47877	364	0,89692	0,09672	10,3
pm	117,37264	35	3,35350	0,05730	6,1
bpm.e	576,25236	1820	0,31662	0,31662	33,8
Toplam	2330,97956	2543			100

Tablo 18'e göre, birey toplam varyans sıralamasında ikinci sırada yer almaktadır ve %29,9'unu açıklamaktadır (kestirilen varyans: 0,28009). Bu durum, öğrenciler (birey, b) arasında performans farklılıkları olduğunu göstermektedir.

Puanlayıcı (p) varyans değeri (0,08150) toplam varyansın %8,7'sini açıklamaktadır. Bu sonuç, puanlayıcılar arasında puanlama düzeylerinin farklılık gösterdiğine işaret etmektedir. Puanlayıcı varyans yüzdesinin benzer çalışmalara göre daha yüksek olmasını, bu çalışmanın önemli bir sonucu olarak görmekteyiz. Nalbantoğlu Yılmaz ve Uzun Başusta (2015) çalışmasına göre puanlayıcı oranı daha fazladır.

Puanlayıcı (öğretim üyesi, p) ve öğrenci bileşeni (bp bileşeni) 0.10040'lük varyans ve %10,7 ile üçüncü en yüksek değere sahiptir. İkili bileşenlerde en yüksek yüzde b x p bileşeninde görülmüştür. Puanlayıcı varyans yüzdesinin benzer çalışmalara göre daha yüksek olmasını, bu çalışmanın önemli bir sonucu olarak görmekteyiz. Nalbantoğlu Yılmaz ve Uzun Başusta'nın (2015) çalışmasına göre puanlayıcı oranı daha fazladır. Birey x puanlayıcı etkileşimini daha ayrıntılı araştıran çalışmaların gerekliliği görülmektedir (Nalbantoğlu Yılmaz ve Uzun Başusta, 2015). Öğrenci, puanlayıcı etkileşimi, ölçülmek istenmeyen puanlayıcı etkisinin, ölçmek istenen öğrenci etkisi ile ortaya çıkan değişkenlik kaynağını temsil etmektedir. %10,7'lik bu sonuç, puanlayıcı etkisinin olduğunu ve öğrenci puanının puanlayıcıdan puanlayıcıya değişebildiğini göstermektedir.

Madde (m) %0,5'lik oranla en düşük varyans yüzdesine sahiptir (kestirilen varyans değeri: 0,00499). Bu performans ölçeğini oluşturan maddeler arasında güçlük bakımından fark olmadığının işaretidir. Madde (görev) varyansının çok düşük olması, hazırlanan görevlerin zorluk derecesinin benzer olduğunu göstermektedir. İyi planlanmış ve zorluk düzeyleri benzer bu tür puanlama ölçeklerinde, eğiticinin eksik yönlerin her görev için ayrı ayrı belirlenmesinde çok önemli avantajı bulunmaktadır.

Puanlayıcı x madde (pm) etkileşiminin varyansı 0,05730 ve toplam varyanstaki yüzdesi %6,1 olarak belirlenmiştir. Ölçülmek istenmeyen madde ile puanlayıcı etkisini göstermektedir. Diğer etkileşimlere göre (bp ve bm) yüzdesinin daha düşük olması, puanlamalar arası tutarlılıkta bir miktar farklılık olmakla birlikte, diğerlerine göre daha az etkiye sahip olduğunun işareti olarak düşünülebilir.

İkili bileşenlerden en düşük yüzdenin puanlayıcı x madde bileşeninde olması, görev için istenen performansta fazla farklılık olmadığını işaret etmektedir. Farklı üniversitelerde görev yapan puanlayıcıların bu çalışmada p x m yüzdesinin daha az olması, görev performans beklentilerinin benzer olduğunu da göstermektedir. Benzer sonuçlar Lafave ve Butterwick (2014) tarafından da belirlenmiştir. Sporcu bantlamasında teknik beceri değerlendirilmesinin yapıldığı bu çalışmada, ikili bileşende en yüksek yüzde puanlayıcı ve katılımcı arasında gösterilmiştir.

Öğrenci, madde etkileşimi (bm bileşeni) 0,09672'lik varyans ve %10,3 ile dördüncü en yüksek değere sahiptir. Öğrenci, madde etkileşimi, ölçülmek istenmeyen madde etkisinin,

ölçmek istenen öğrenci etkisi ile ortaya çıkan değişkenlik kaynağını temsil etmektedir. %10,3'lük sonuçla bu bileşen, maddelerin güçlük seviyesinin öğrenciden öğrenciye bir miktar farklılık gösterdiği şeklinde açıklanabilir.

Birey, puanlayıcı ve madde (b x p x m) bileşenini (artık bileşeni), %33,8'lik varyans yüzdesi ve 0,31662'lik varyans ile en yüksek orana sahiptir. Bu değer, üç bileşenin (b, p ve m) ortak etkileşiminden ve/veya ölçme hatalarından kaynaklanmaktadır. Çalışmamızda, bu yüzdenin en fazla olması sonucunda, kalan (artık) bileşenin yüksek olduğu ve başka değişkenlik kaynaklarının da olabileceği söylenebilir.

Çalışmada elde edilen ve tüm maddeleri içeren G katsayısı bağıl hata kaynakları için 0,89 (G göreceli, *relative*) ve mutlak hata kaynakları için ise 0,85 (G mutlak, *absolute*) olarak bulunmuştur.

4.7. Dereceleme Ölçeği puanları için K çalışması

Genellenebilirlik çalışmasında kullanılan verilerle hesaplanan varyans değerleri K çalışmasında elde edilecek kararlar için kullanılmaktadır. K çalışması için, çalışmamızda birey dışındaki puanlayıcı ve madde sayılarının azaltılıp artırılması ile göreli ve mutlak kararlar için sırasıyla G ve Phi katsayılarının kestirilmesine izin verir. Bu amaçla çalışma modelimizde yer alan yüzeylerin en uygun sayısını farklı durumlarda kestirebilmek için K çalışması yürütülmüştür.

Puanlayıcı sayısı sırasıyla, 2, 4, 8 ve madde sayısı da sırasıyla 6, 10 olarak alınmıştır. Elde edilen veriler Tablo 19 ve 20'de verilmiştir.

Tablo 15. b x m x p Desenli K çalışması analiz sonuçları

		Opsiyon				
	G-çalışması	**1**	**2**	**3**	**4**	**5**
B (birey)	53	53	53	53	53	53
P (puanlayıcı)	6	2	4	8	6	6
M (madde)	8	8	8	8	6	10
Observ.	2544	848	1696	3392	1908	3180
Coef_G rel.	0.88774	0.77338	0.85609	0.90446	0.87056	0.89838
Coef_G abs.	0.84643	0.68798	0.80035	0.87152	0.82930	0.85705
Rel. Err. Var.	0.03542	0.08208	0.04708	0.02959	0.04165	0.03168
Rel. Std. Err. of M.	0.18820	0.28649	0.21699	0.17201	0.20408	0.17799
Abs. Err. Var.	0.05082	0.12703	0.06987	0.04129	0.05765	0.04672
Abs. Std. Err. of M.	0.22543	0.35641	0.26433	0.20321	0.24011	0.21615

Tablo 20. b x m x p Desenli K çalışması analizinin puanlayıcı ve madde sayısına göre sıralı sonuçları

n_b	n_p	n_m	G (Coef_G rel)	Phi (Coef_G abs.)	$\sigma^2(\delta)$ (Rel. Err. Var.)	$\sigma^2(\Delta)$ (Abs. Err. Var.)
53	2	8	0.77338	0.68798	0.08208	0.12703
53	4	8	0.85609	0.80035	0.04708	0.06987
53	8	8	0.90446	0.87152	0.02959	0.04129
53	6	6	0.87056	0.82930	0.04165	0.05765
53	**6**	**8**	**0.88774**	**0.84643**	**0.03542**	**0.05082**
53	6	10	0.89838	0.85705	0.03168	0.04672

Koyu satır, çalışmanın normal desenini göstermektedir.

İlk durumda, madde sayısı sabit, puanlayıcı sayısı değiştirildiğinde, puanlayıcı sayısı arttıkça göreceli ve mutlak hata varyanslarının azaldığı, G ve Phi katsayılarının ise arttığı görülmektedir. Benzer durum, madde sayısı için de geçerlidir. Puanlayıcı sayısının sabit, madde sayısının ise değiştirildiği durum için madde sayısı arttıkça göreli ve mutlak hata varyanslarının azaldığı, G ve Phi katsayılarının ise yine arttığı görülmektedir. Her iki durumdaki değişiklikler incelendiğinde, puanlayıcı sayısındaki koşul artışının, madde sayısının artışındaki değişiklikten daha fazla olduğu görülmektedir. Bu durum, puanlayıcılar arasındaki birey değerlendirme farklılıklarının maddelerden daha fazla etkiye sahip olduğunun işareti olarak kabul edilebilir.

4.8. Klasik Test Kuramı (KTK) ve Genellenebilirlik Kuramı (G-Kuramı) sonuçlarının karşılaştırılması

Dereceleme ölçeğinden elde edilen sonuçların KTK ve b x p x m modelli G-Kuramına göre elde edilen sonuçlar aşağıdaki tabloda gösterilmiştir (Tablo 21).

Tablo 16. Dereceleme ölçeğine göre elde edilen sonuçların KTK ve G-Kuramına göre sonuçları

Klasik Test Kuramı:	
Maddelerin puanlayıcılar arası güvenirliği:	ICC=0,766-0,919
Ölçek toplam puanının puanlayıcılar arası güvenirliği:	ICC=0,926
Ölçeğin iç tutarlılığı	Alfa=0,853
Ölçek toplam puanının puanlayıcılar arası ilişkisi:	r=0,681-0,910 (p<0,001)
Genellenebilirlik Kuramı	
b x p x m	Varyans yüzdeleri:
b	%29,9
p	%8,7
m	%0,5
bp	%10,7
bm	%10,3
pm	%6,1
bpm,e	%33,8
G Göreceli	0,89
G Mutlak	0,85

r: Pearson korelasyon katsayısı. ICC: Sınıf içi korelasyon katsayısı.

Klasik test kuramına göre incelendiğinde çalışmada puanlama yapan 6 puanlayıcı arasındaki güvenirlik 0,926 olarak oldukça yüksek (mükemmel) bulunmuştur. Benzer sonuçlar her madde bazında da yüksek çıkmıştır. Ancak yapılan 15 t testi karşılaştırmasında (bağımlı örneklerde t testi) 1.-4., 1.-6. ve 4.-6. puanlayıcı dışındaki (p>0,05) 12

karşılaştırmada tüm puanlayıcı puanlamaları arasında fark bulunmuştur ($p<0,05$). Çalışmada, altı puanlayıcıya ait Cronbach Alfa katsayılarının ortancası hesaplanmıştır (0,853). Elde edilen değer, G katsayısına benzer bulunmuştur. G Kuramına göre incelendiğinde, klasik kuramdaki puanlayıcılar arası mükemmel güvenirliğe rağmen, puanlayıcıların varyans yüzdesinin yüksek olduğu (%8,7), puanlayıcı öğrenci bileşeninde de varyans yüzdesinin %10,7 olduğu gözlenmiştir. Ayrıca kalan (artık) bileşenin yüksek olması (%33,8), başka değişkenlik kaynaklarının da olabileceğine işaret etmektedir. Özellikle puanlayıcıların dereceleme ölçeği puanlama sistematiğine dair deneyimlerinin olup olmadığı, değerlendirilen öğrenci sayısının yorgunluk etkisi gibi faktörlere ilişkin daha ayrıntılı çalışmaların gerekliliğine işaret etmektedir.

BEŞİNCİ BÖLÜM

SONUÇ ve ÖNERİLER

Sonuç ve Öneriler bölümünde bir önceki bölümde ulaşılan sonuçların özetine ve çalışma ve uygulamalara yönelik önerilere yer verilmiştir.

5.1. Sonuçlar

5.1.1. Test-tekrar test sonuçları

Dereceleme ölçeği ile genel izlenim puanlamasında test-tekrar test güvenirliğinin incelenmesinde toplam puanlar arası güvenirlik her iki puanlama türünde de sınıf içi korelasyon analizine göre mükemmel bulunmuştur. Dereceleme ölçeği maddelerinin, maddeler düzeyinde sınıf içi korelasyon katsayıları tüm maddeler için mükemmel olarak bulunmuştur. Puanlayıcıların farklı zamanlarda aynı öğrenciyi değerlendirirken puanlamaları arasındaki uyumunun yüksek olduğunu göstermektedir. Maddeler için de benzer şekilde, aynı değerlendiricinin farklı zamanda her madde için yüksek oranda uyum gösterdiği sonucuna varmamızı sağlamıştır.

5.1.2. Dereceleme ölçeğinin iç tutarlılığı

Dereceleme ölçeğinin iç tutarlılığı tüm puanlayıcılar için 0,80'in üzerinde bulunmuştur. İç tutarlılığın 6 puanlayıcıda da yüksek olması, geliştirilen dereceleme ölçeğini oluşturan maddelerdeki tutarlılığın yeterli olduğunu ve fizyoterapi ve rehabilitasyon eğitiminin diğer performans durum belirlemelerinde de kullanılabileceğini işaret etmektedir.

5.1.3. Dereceleme Ölçeği ile Genel İzlenim puanlamalarının puanlayıcılar arası karşılaştırılması

Çalışmaya katılan 6 puanlayıcı arasındaki güvenirlik sınıf içi korelasyon katsayısı ve Pearson korelasyon analizi ile incelenmiştir.

Dereceleme ölçeğinde maddelerin her biri için güvenirlik iyi ve mükemmel arasında değişmekteydi. Toplam puan için ise güvenirlik mükemmel bulunmuştur. Genel izlenim puanlamasında, aynı puanlayıcıların güvenirlikleri dereceli ölçekte olduğu gibi mükemmel

çıkmıştır. Puanlayıcılar arasındaki ilişki yüksek bulunmuştur. Puanlayıcıların verdiği puanlar t testi ile karşılaştırıldığında ise çoğunlukla her iki puanlama türünde de (dereceleme ölçeği ve genel izlenim) arada fark olduğu belirlendi. Puanlayıcılar arası yüksek güvenirliğe rağmen puanlarda farkın olması, puanlayıcıların puan verme süreçlerini etkileyen başka değişkenlerin olabileceğini göstermektedir.

5.1.4. Dereceleme Ölçeğinin Genellenebilirlik Kuramına göre incelenmesi

Çalışmada 6 puanlayıcıdan 53 öğrenci üzerinde 8 farklı madde ile puanlanan veriler genellenebilirlik kuramı ile b x p x m desenine göre analiz edilmiştir.

Öğrenci (birey) grubu toplam varyans sıralamasında ikinci sırada yer almıştır. Bu durum, öğrenciler (birey, b) arasında farklılıklar olduğunu göstermektedir. Öğrencilerin pratik performans farklılıklarının bu seviyede olması kabul edilebilir düzeydedir. Puanlayıcı varyans yüzdesinin yüksek olmasını, bu çalışmanın önemli bir sonucu olarak görmekteyiz. İkili bileşenlerde en yüksek yüzde birey x puanlayıcı bileşeninde görülmüştür. Birey x puanlayıcı etkileşimine etki edebilecek etmenleri araştıran çalışmaların yapılmasının yararlı olacağı görüşündeyiz.

Puanlayıcıların belirli bir düzeye getirilmesi, puanlayıcı eğitiminin önemli bir bileşenidir. Pratik bir uygulamada performans değerlendirmesinde, öznel puanlayıcı etkilerini azaltabilmek amacıyla, ortak ve görüş birliği içeren çalışmalar yapılmalıdır.

Madde (görev) varyansının çok düşük olması, hazırlanan görevlerin zorluk derecesinin benzer olduğunu göstermektedir. İkili bileşenlerden en düşük yüzdenin puanlayıcı x madde bileşeninde olması, görev için istenen performansta fazla farklılık olmadığını göstermektedir. Puanlayıcıların farklı üniversitelerden olduğu bu çalışmada, puanlayıcı x madde bileşen yüzdesinin daha az olması, görev performans beklentilerinin benzer olduğunu da göstermektedir.

Artık bileşenin en yüksek oranda olması (%33,8), olabilecek diğer nedenler konusunda daha ayrıntılı çalışmaların yapılması gerektiğini ortaya koymaktadır. En yüksek yüzdede olmasına rağmen, diğer çalışmalara göre bu oranın düşük olmasında üçlü çapraz modelin ve

puanlamada evet-hayır şeklinde değil, dereceli puanlamanın (0, 1, 2, 3) bu oranı düşük tuttuğunu düşünmekteyiz.

5.1.5. Dereceleme Ölçeğinin Genellenebilirlik Kuramında Karar (K) çalışması

İlk durumda, madde sayısı sabit, puanlayıcı sayısı değiştirildiğinde, puanlayıcı sayısı arttıkça göreceli ve mutlak hata varyanslarının azaldığı, G ve Phi katsayılarının ise arttığı görülmektedir. Benzer durum, madde sayısı için de geçerlidir. Puanlayıcı sayısının sabit, madde sayısının ise değiştirildiği durum için madde sayısı arttıkça göreli ve mutlak hata varyanslarının azaldığı, G ve Phi katsayılarının ise yine arttığı görülmektedir. Her iki durumdaki değişiklikler incelendiğinde, puanlayıcı sayısındaki koşul artışının, madde sayısının artışındaki değişiklikten daha fazla olduğu görülmektedir. Bu durum, puanlayıcılar arasındaki birey değerlendirme farklılıklarının maddelerden daha fazla etkiye sahip olduğunun işareti olarak kabul edilebilir.

5.1.6. Klasik Test Kuramı (KTK) ve Genellenebilirlik Kuramı (G-Kuramı) sonuçlarının karşılaştırılması

Çalışmada Klasik Test Kuramına göre puanlayıcılar arası güvenirlik sınıf içi korelasyon katsayısı ile incelenmiştir. Her iki değerlendirme yöntemi için de mükemmel bulunmuştur (ICC: 0,926 ve 0,943). Yüksek bulunan bu değere rağmen, puanlayıcılar arasında her iki puanlama türüne göre farklılıklar bulunmuştur. Genellenebilirlik Kuramına göre incelendiğinde ise, çok daha ayrıntılı sonuçlar ortaya konulabilmiştir. Puanlayıcılar arası farkların ve puanlayıcı öğrenci etkileşimlerine yönelik daha ayrıntılı yeni çalışmalara ihtiyaç olduğu görülmektedir.

Genellenebilirlik Kuramının fizyoterapi ve rehabilitasyon eğitiminde performans değerlendirilmesinde ilk kez kullanılmış olması nedeniyle, çalışmamızın sonraki çalışmalar için referans olacağı görüşündeyiz.

5.2. Öneriler

1. Fizyoterapi ve rehabilitasyon alanında, analitik performans değerlendirme yöntemlerinin daha yaygın ve daha farklı alanlarda çalışılması gerektiğini düşünmekteyiz.

2. Farklı pratik performans türlerinin ve görevlerinin aynı anda karşılaştırılacağı çalışmalar konuya katkı sağlayacaktır. Nitekim, Lafave ve Butterwick'in (2014) çalışmasında farklı vücut bölgesi için yapılan performans uygulamasının varyans yüzdesi %67,84 olarak gösterilmiştir.

3. Performans göstergelerinin hangi ölçütlere göre değerlendirileceğini ve performanslarının hangi puana denk geldiğini gösteren puanlama aracı ile (rubrik) yapılacak Genellenebilirlik Kuramı çalışmalarının yararlı olacağı görüşündeyiz.

4. Temel pratik bilgi (örneğin anatomi dersi), öğrenci üzerinde pratik uygulama (örneğin elektroterapi) ve hasta üzerinde pratik (örneğin klinik çalışma) performansların denendiği çalışmaların gerektiğini düşünmekteyiz.

5. Genellenebilirlik ve Karar çalışmalarında analitik derecelendirmede görev özelliğinin performansa etkisinin az olması ve görevlerin her birinde eksikliklerin görülebilmesi nedeniyle, diğer değerlendirme alanlarında da kullanılabileceğini düşünmekteyiz.

6. Performans değerlendiricilerin veya puanlayıcıların, değerlendirme aşamasında öğrenci etkilenimlerini ve nedenlerini araştıran çalışmaların yapılması yararlı olacaktır.

7. Faklı üniversitelerde bu tür pratik performans değerlendirme çalışmalarının yapılması ve uygulamaya konmasının fizyoterapi ve rehabilitasyon eğitimindeki standardizasyon açısından yararlı olacağını düşünmekteyiz.

8. Hasta veya hasta rolündeki model öğrenci değişikliğinin etkisi, farklı vücut bölgelerinde aynı performansın değerlendirilmesini araştıran çalışmalar denenmelidir.

9. Özellikle puanlayıcıların analitik puanlama sistematiğine dair deneyimlerinin olup olmadığı, değerlendirilen öğrenci sayısının yorgunluğa etkisi gibi faktörlere ilişkin daha ayrıntılı çalışmalar gerçekleştirilebilir.

10. Çalışmamızda kullanılan b x p x m deseni dışında farklı desenlerde çalışmalar yapılabilir.

11. Dereceli puanlamada derece sayısındaki değişikliklerin etkisini araştıran çalışmalar yapılabilir.

12. Öğrencinin kendini değerlendirmesi sonuçları ile puanlayıcının analitik değerlendirme sonuçlarını karşılaştıran ve irdeleyen çalışmalar planlanabilir.

13. Fizyoterapi ve rehabilitasyon eğitimi veren kurumlarda performans değerlendirme standartlarına yönelik ortak çalışmalar yapılmalıdır.

14. Fizyoterapi ve rehabilitasyon eğitimi veren kurumlarda performans değerlendirme standartlarının yaygınlaştırılmasına dair mesleki eğitim platformlarında çalışmalar planlanmalıdır.

15. Bu çalışmada performans değerlendirilmesi olarak fiziksel bir uygulama alınmıştır. Psikolojik özellikler ile fiziksel performans değerlendirmelerini karşılaştırılan çalışmalar planlanabilir.

KAYNAKÇA

Adamson, K., Kardong-Edgren, S., & Willhaus, J. (2012). An updated review of published simulation evaluation instruments. *Clinical Simulation in Nursing, 9*(9), e393–e400.

Atılgan, H. (2005). Genellenebilirlik Kuramı ve Puanlayıcılar Arası Güvenirlik İçin Örnek Bir Uygulama. *Eğitim Bilimleri ve Uygulama, 4*(7), 95-108.

Atılgan, H. (2019). *Genellenebilirlik Kuramı ve Uygulaması.* Ankara: Anı Yayıncılık.

Başaranoğlu, G. (2018). Yapılandırılmış objektif klinik sınav: Bezmialem Vakıf Üniversitesi, Sağlık Hizmetleri meslek Yüksekokulu, Anestezi Programı Deneyimi. *Yükseköğretim ve Bilim Dergisi, 8*(2), 388-391.

Baykul, Y. (2015). *Eğitimde ve Psikolojide Ölçme Klasik Test Teorisi ve Uygulaması. 3. Baskı.* Ankara: Pegem Akademi.

Bekiroğlu, F. (2008). Performansa Dayalı Ölçümler: Teori ve Uygulama. *Türk Fen Eğitim Dergisi, 5*(1), 113-131.

Birenbaum, M. (1994). Toward adaptive assessment-the student's angle. *Studies in Educational Evaluation*, 20, 239-255.

Bloch, R., & Norman, G. (2012). Generalizability theory for the perplexed: A practical introduction and guide: AMEE Guide No. 68. *Medical Teacher, 34*(11), 960-992.

Brennan, R. L. (2001). *Generalizability Theory.* New York: Springer Verlag.

Brualdi, A. (1998). Implementing Performance Assessment in the Classroom. *Practical Assessment, Research, and Evaluation, 6* (2), 1-3.

Büyükkıdık, S. (2012). Problem Çözme Becerisinin Değerlendirilmesinde Puanlayıcılar Arası Güvenirliğin Klasik Test Kuramı ve Genellenebilirlik Kuramına Göre Karşılaştırılması. *Yayımlanmış Yüksek Lisans Tezi.* Ankara: Hacettepe Üniversitesi.

Büyükkıdık, S., & Anıl, D. (2015). Performansa Dayalı Durum Belirlemede Güvenirliğin Genellenebilirlik Kuramında Farklı Desenlerle İncelenmesi. *Eğitim ve Bilim Dergisi, 40*(177), 285-296.

Büyüköztürk, Ş. (2007). Performansa Dayalı Durum Belirleme nedir? *İlköğretmen Dergisi, 8*, 28-32.-.

Büyüköztürk, Ş., E., K. Ç., Akgün, Ö., Ş., K., & Demirel, F. (2019). *Eğitimde Bilimsel Araştırma Yöntemleri* (27 b.). Ankara: Pegem Akademi.

Carroll University Physical Therapy Program. (2013). January 5, 2020 tarihinde https://my.carrollu.edu/ICS/icsfs/PTH414.pdf?target=ceb8fd96-77e8-447b-8afb-819415e78a2e adresinden alındı

Chong, D. Y., Tam, B., Yau, S. Y., & & Wong, A. Y. (2020). Learning to prescribe and instruct exercise in physiotherapy education through authentic continuous assessment and rubrics. *BMC medical education, 20*(1), 1-11.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. Journal of Applied Psychology, 78(1): 98.

Cronbach, L. J. (1963). Theory of generalizability: A liberalization of reliability theory. *British Journal of Statistical Psychology, 16*(2), 137-163.

Crocker, L., Algina, J. (1986). Introduction Classical and Modern Test Theory. USA: CBS College Publishing Company.

Cross, V. (1983). Student evaluation and assessment in clinical locations Physiotherapy. *Physiotherapy, 69*(9), 304-308.

Cross, V. and Hicks, C. (1997). What Do Clinical Educators Look for in Physiotherapy Students? Physiotherapy, 83 (5): 249-260.

Dalton M, Davidson M, Keating J. The Assessment of Physiother-apy Practice (APP) is a valid measure of professional competence ofphysiotherapy students: a cross-sectional study with Rasch analysis. J Physiother 2011;57(4):239–46.

Daniels, V. J., & Pugh, D. (2018). Twelve tips for developing an OSCE that measures what you want. *Medical teacher, 40*(12), 1208-1213.

Deliceoğlu, G. (2009). Futbol yetilerine ilişkin dereceleme ölçeğinin genellenebilirlik ve klasik test kuramına dayalı güvenirliklerinin karşılaştırılması. Doktora Tezi, Ankara: Ankara Üniversitesi.

Denat, Y., & Tuğrul, E. (2012). Klinik beceri performanslarını değerlendirmede bir yöntem: objektif yapılandırılmış klinik sınavlar. *Hemşirelikte Eğitim ve Araştırma, 9*(3), 53-59.

Dochy, F., and McDowell, L. (1997). Assessment as a tool for learning. *Studies in Educational Evaluation*, 23(4), 279-298.

Doğan, C. D., & Anadol, H. Ö. (2017). Genellenebilirlik kuramında tümüyle çaprazlanmış ve maddelerin puanlayıcılara yuvalandığı desenlerin karşılaştırılması. *Kastamonu Eğitim dergisi, 25*(1), 361-372.

Doğan, C. D., & Yosmaoğlu, H. B. (2015). The effects of the analytical rurics on the objectivity in physiotherapy practical examination. *Türkiye Klinikleri Journal of Sports Science, 7*(1), 9-15.

Erkuş, A., Sünbül, Ö., Ömür Sünbül, S., Aşiret, S., & Yormaz, S. (2017). *Psikolojide ölçme ve ölçek geliştirme.* Ankara: Pegem Akademi.

Gatti, A. A., Stratford, P. W., Brisson, N. M., & Maly, M. R. (2020). How to Optimize Measurement Protocols: An Example of Assessing Measurement Reliability Using Generalizability Theory. *Physiother Canada, 72*(2), 112-121.

Goodrich, H. (1997). Understanding rubrics. *Educational Leadership, 54*(4), 14-17.

Güler, N., Kaya Uyanık, G., & Taşdelen Teker, G. (2012). *Genellenebilirlik Kuramı.* Ankara: Pegem Akademi.

Gwet, K. L. (2020). *Her Yönüyle Puanlayıcılar Arası Güvenirlik Rehberi .* (i. Karakaya, & H. Yürekli, Çev.) Ankara: Pegem Akademi.

Haider, I., Badshah, A., Khan, A. R., & Ullah, A. (2018). Fatigue level of examiners during objective structured clinical examination (OSCE). *Journal of Medical Sciences, 26*(3), 207-210.

Harden, R. M., Stevenson, M. D., & Wilson, G. M. (1975). Assessment of clinical competence using objective structured examination. *British Medical Journal, 1*(5955), 447-451.

Holey, L. A. (1993). A new way to assess practical physiotherapy skills. *Medical Teacher, 15*(4), 379-386.

Hrachovy J, Clopton N, Baggett K, Garber T, Cantwell J, SchreiberJ. Use of the Blue MACS: acceptance by clinical instructors and self-reports of adherence. Phys Ther 2000;80(7):652–61.

Irwin, J. E. (2019). 2nd year doctor of physical therapy students less proficient at goniometry than the reported norm. *Int Phys Med Rehab J. , 4*(4), 161-164.

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of chiropractic medicine, 15*(2), 155–163.

Köse, A. (2014). Ölçmede Güvenirlik. R. N. Demirtaşlı içinde, Eğitimde ölçme ve değerlendirme (s. 86-109). Ankara: Edge Akademi.

Kreiter, C. a. (2020). Generalizability Theory's Role in Validity Research: Innovative Applications in Health Science Education. *Health Professions Education, 6*(2), 282-290.

Kurtz, S., Silverman, J., & Draper, J. (1998). *Teaching and Learning Communication Skills in Medicine.* Oxford: Radcliffe Medical Press.

Kutlu, Ö., Doğan, C., & Karakaya, İ. (2017). *Ölçme ve Değerlendirme: Performansa ve Portfolyoya Dayalı Durum Belirleme, Beşinci Baskı.* Ankara: Pegem Akademi.

Lafave, M., & Butterwick, D. (2014). A generalizability theory study of athletic taping using the Technical Skill Assessment Instrument. *Journal of athletic training, 49*(3), 368-372.

Mariani, B., & Doolen, J. (2016). Nursing simulation research: What are the perceived gaps? *Clinical Simulation in Nursing, 12*, 30-36.

McGaghie, W. C., Butter, J., & Kaye, M. (2009). Observational Assessment. R. Y. Steven M. Downing içinde, *Assessment in Health Professions Education* (s. 185-216). Routledge.

McGinty, S. M. (2000). Case-method teaching: an overview of the pedagogy and rationale for its use in physical therapy education. *Journal of Physical Therapy Education, 14*(1), 48.

Medley, D. (1984). Teacher Competency testing and teacher education. L. Katz, & J. Raths içinde, *Advances in teacher education* (Cilt 1, s. 51-94). Norwood, NJ: Ablex.

Miller, G. (1990). The assessment of clinical skills/competence/performance. *Academic medicine, 65*(9), S63-7.

Monteiro, S., Sullivan, G. M., & Chan, T. M. (2019). Generalizability Theory Made Simple(r): An Introductory Primer to G-Studies. *Journal of Graduate Medical Education, 11*(4), 365–370.

Myezwa H, Roos R, Mudzi W, Potterton J, Watt B, Maleka D, Mtshali S, Stewart A. (2013). Inter-examiner reliability when using the objective structured practical examination (OSPE) mark sheet for physiotherapy practical examinations. *South African Journal of Physiotherapy*, Special Edition(4), 21-28.

Nalbantoğlu Yılmaz, F., Uzun Başusta, B. (2015). Genellenebilirlik kuramıyla dikiş atma ve alma becerileri istasyonu güvenirliğinin değerlendirilmesi. Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi. 6(1): 107-116.)

Nayar, U., Malik, S. L., & Bijlani, R. L. (1986). Objective structured practical examination: a new concept in assessment of laboratory exercises in preclinical sciences. *Medical education, 20*(3), 204-209.

Norkin, C. C., & White, D. J. (2016). *Measurement of Joint Motion: A Guide to Goniometry* (5 b.). Philadelphia: F. A. Davis Company.

O'Brien, J., Thompson, M. S., & Hagler, D. (2019). Using generalizability theory to inform optimal design for a nursing performance assessment. *Evaluation & the Health Professions*, 297-327.

O'Connor, A., McGarr, O., Cantillon, P., McCurtin, A., & Clifford, A. (2018). Clinical performance assessment tools in physiotherapy practice education: a systematic review. *Physiotherapy, 104*(1), 46-53.

Olivier, B., Naidoo, V., Humphries, C., Godlwana, L., Romm, M., Ntsiea, V., . . . Stewart, A. (2013). Inter-examiner reliability when using the Objective Structured Practical Examination (OSPE) mark sheet for physiotherapy practical. *South African Journal Physiotherapy, 69*(4), 21-28.

Özçelik, D. A. (2013). Okullarda Ölçme ve Değerlendirme. Öğretmen El Kitabı. 2. Baskı. Ankara: Pegem Akademi.

Öztürk, M. E. (2011). Voleybol becerileri gözlem formu ile elde edilen puanların genellenebilirlik ve klasik test kuramına göre karşılaştırılması. *Yüksek Lisans Tezi*, Ankara: Hacettepe Üniversitesi.

Palm, T. (2008). Performance Assessment and Authentic Assessment: A Conceptual Analysis of the Literature. *Practical Assessment, Research & Evaluation, 13*(4), 1-11.

Portney, L. G., & Watkins, M. (2015). *Foundations of Clinical Research: Applications to Practice* (3 b.). Philadelphia: F. A. Davis Company.

Preus, R. A. (2013). Using generalizability theory to develop clinical assessment protocols. *Phys therapy*, *93*(4), 562-569.

Prion, S. K., Gilbert, G. E., & Haerling, K. A. (2016). Generalizability theory: An introduction with application to simulation evaluation. *Clinical Simulation in Nursing, 12*, 546–554.

Ribeiro, A. M., Ferla, A. A., & Amorim, J. S. (2019). Objective structured clinical examination in physiotherapy teaching: a systematic review. *Fisioterapia em Movimento, 32*, e003214.

Roach KE, Frost JS, Francis NJ, Giles S, Nordrum JT, Delitto A. Valida-tion of the Revised Physical Therapist Clinical Performance Instrument(PT CPI): version 2006. Phys Ther 2012;92(3):416–28.

Sakurai, H., Kanada, Y., Sugiura, Y., Motoya, I., Wada, Y., Yamada, M., . . . Okanishi, T. (2014, September). OSCE-based Clinical Skill Education for Physical and Occupational Therapists . *Journal of physical therapy science, 26*(9), 1387-1397.

Sattelmayer, M., Hilfiker, R., & Baer, G. (2017). A systematic review of assessments for procedural skills in physiotherapy education. *International Journal of Health Professions, 4*(1), 53-65.

Shavelson, J. R., & Webb, N. M. (2006). Generalizability theory. J. Green, G. Camill, & P. e. Elmore içinde, *Handbook of complementary methods in education research* (s. 309-322). Mahwah: Lawrence Erlbaum Associates Publishers.

Suen, H. K., & Lei, P. (2007). Classical Versus Generalizability Theory of Measurement. *Educational Measurement, 4*, 1-13.

Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education,* 48: 1273-1296.

Turgut, M. F., Baykul, Y. Eğitimde Ölçme ve Değerlendirme. 8. Baskı, Ankara: Pegem Akademi.

University of Minnesota, Department of Rehabilitation, Division of Physical Therapy. (2019). January 1, 2020 tarihinde https://canvas.umn.edu/courses/141800/files/8106212 adresinden alındı

Usherwood, T., Challis, M., Joesbury, H., & & Hannay, D. (1995). Competence-based summative assessment of a student-directed course: involvement of key stakeholders. *Medical Education, 29*(2), 144-149.

Uzun, N. B., Aktaş, M., Aşiret, S., & Yorulmaz, S. (2018). Using Generalizability Theory to Assess the Score Reliability of Communication Skills of Dentistry Students. *Asian Journal of Education and Training, 4*(2), 85-90.

van de Watering, G., Gijbels, D., Dochy., F. van der Rijt, J. (2008). Students' assessment preferences, perceptions of assessment and their relationships to study results. *High Educ*. 56:645-658.

Vendrely, A. (2002). Student assessment methods in physical therapy education: an overview and literature review. *Journal of Physical Therapy Education, 16*(2), 64-69.

World Confederation for Physical Therapy. WCPT guideline forphysical therapist professional entry level education. London, UK: WCPT; 2011. (tarih yok). January 6, 2021 tarihinde World Physiotherapy: www.wcpt.org/guidelines/entry-level-education adresinden alındı

Yaqinuddin, A., Zafar, M., Ikram, M. F., & Ganguly, P. (2013). What is an objective structured practical examination in anatomy? *Anatomical sciences education, 6*(2), 125-133.

Yıldırım Seheryeli, M. (2018). Yazılı anlatım becerisi puanlama anahtarının güvenirliğinin klasik test, genellenebilirlik ve madde tepki kuramlarına göre incelenmesi. *Yayımlanmış Yüksek Lisans Tezi*. Ankara: Gazi Üniversitesi.

Yıldıztekin, B. (2014). Klasik Test Kuramı ve Genellenebilirlik Kuramından Puanlayıcılar Arası Tutarlılığın Farklı Yöntemlere Göre Karşılaştırılması. *Yayımlanmış Yüksek Lisans Tezi*. Ankara: Hacettepe Üniversitesi.

Yılmaz, N. (2012). Genellenebilirlik Kuramında Dengelenmiş ve Dengelenmemiş Desenlerin Karşılaştırılması -İntramusküler Enjeksiyon Yapma İstasyonu Verileri Üzerinde Bir Uygulama. *Yayınlanmış Doktora Tezi*. Ankara: Hacettepe Üniversites.

Yılmaz, N., & Başbaşa, N. B. (2015). Assessment of sewing and picking skills station reliability with generability theory. *Journal of Measurement and Evaluation in Education and Psychology, 6*(1), 107-116.

Yılmaz, N., & Gelbal, S. (2011). İletişim becerileri istasyonu örneğinde genellenebilirlik kuramıyla farklı desenlerin karşılaştırılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 41,* 509-518.

Yılmaz, N., & Tavsancıl, E. (2014). Comparison of balanced and unbalanced patterns in generalizable theory with intramuscular injection station data. *Education and Science, 39*(175), 285-295.

EK 1. Çalışmanın Etik Onayı.

T.C.
HASAN KALYONCU ÜNİVERSİTESİ
Sağlık Bilimleri Fakültesi
Girişimsel Olmayan Araştırmalar Etik Kurul Kararı

Karar No : 2020/034
Karar Tarihi : 28.05.2020

Sayın Prof. Dr. Yavuz YAKUT,

"Fizyoterapi ve Rehabilitasyon Öğrencilerinde Eklem Hareket Ölçüm Becerilerinin Pratik Değerlendirilmesinin Genellenebilirlik Kuramına Göre İncelenmesi" konulu çalışmanızın girişimsel olmayan araştırmalar etik kurul kararı uyarınca uygun olduğuna;
Oy birliği ile karar verilmiştir.

Prof. Dr. Zerrin PELİN
Başkan

Prof. Dr. Yasemin BEYHAN
Üye

Prof. Dr. S. Mine YURTTAGÜL
Üye

Prof. Dr. Nermin OLGUN
Üye

Prof. Dr. Kezban BAYRAMLAR
Üye

(Sorumlu Araştırmacı Olduğundan Katılmadı)
Prof. Dr. Yavuz YAKUT
Üye

Prof. Dr. Ayla YAVA
Üye

Prof. Dr. Tülay ORTABAĞ
Üye

ASLIGIBIDIR