



HACETTEPE ÜNİVERSİTESİ EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı

Eğitimde Ölçme ve Değerlendirme Programı

PERFORMANS DEĞERLENDİRMEDE PUANLAYICILAR ARASI UYUM İLE GÖZ İZLEME SONUÇLARI ARASINDAKİ İLİŞKİNİN İNCELENMESİ

Mine DEMİRBAŞ

Doktora Tezi

Ankara, 2023

Liderlik, arařtırma, inovasyon, kaliteli eđitim ve deđiřim ile

Daha ileriye... En İyiyeye...



Eğitim Bilimleri Ana Bilim Dalı
Eğitimde Ölçme ve Değerlendirme Programı

PERFORMANS DEĞERLENDİRMEDE PUANLAYICILAR ARASI UYUM İLE GÖZ İZLEME
SONUÇLARI ARASINDAKİ İLİŞKİNİN İNCELENMESİ

INVESTIGATION OF INTERRATER AGREEMENT AND EYE TRACKING IN
PERFORMANCE EVALUATION

Mine DEMİRBAŞ

Doktora Tezi

Ankara, 2023

Kabul ve Onay

Eđitim Bilimleri Enstitüsü M¼d¼rl¼đ¼ne,

Mine Demirbař'ın hazırladıđı "Performans Deđerlendirmede Puanlayıcılar Arası Uyum İle G¼z İzleme Sonuçları Arasındaki İliřkinin İncelenmesi " bařlıklı bu çalıřma j¼rimiz tarafından **Eđitim Bilimleri Ana Bilim Dalı, Eđitimde Ölçme ve Deđerlendirme Bilim Dalında Doktora Tezi** olarak kabul edilmiřtir.

J¼ri Bařkanı Prof.Dr. Selahattin GELBAL

J¼ri Üyesi (Danıřman) Prof.Dr.Nuri DOđAN

J¼ri Üyesi Prof.Dr.H¼lyya KELECİOđLU

J¼ri Üyesi Prof.Dr.İsmail KARAKAYA

J¼ri Üyesi Prof.Dr.Hakan Yavuz ATAR

Bu tez Hacettepe Üniversitesi Lisans¼st¼ Eđitim, Öğretim ve Sınav Y¼netmeliđi'nin ilgili maddeleri uyarınca yukarıdaki j¼ri üyeleri tarafından / / tarihinde uygun g¼r¼lm¼ř ve Enstit¼ Y¼netim Kurulunca / / tarihi itibarıyla kabul edilmiřtir.

Prof. Dr. Selahattin GELBAL
Eđitim Bilimleri Enstitüsü M¼d¼r¼

Öz

Bu arařtırmada performans deęerlendirme srecinde puanlayıcılar arası uyum, uyum katsayıları, genellenebilirlik ve ok yzeyli Rasch ile arařtırılmıřtır. Bu ama doęrultusunda oęrencilerin sunum yapma performansını deęerlendirmek zere aynı puanlayıcıların sınıf ii performans esnasında verdikleri puanlar ve gz izleme teknięi ile ekran zerinden verdikleri puanlar arasındaki iliřkiler incelenmiřtir. Sınıf ii deęerlendirme ve gz izleme teknięi ile ekran zerinden yapılan deęerlendirme sonucu elde edilen puanlar, klasik test kuramında yer alan puanlayıcılar arası uyum istatistikleri, genellenebilirlik kuramı ve ok yzeyli Rasch modeline gre analiz edilmiřtir. Arařtırma sonucunda, idealizm ve realizm sunumları iin puanlayıcılar arası uyumun iyi olduęu bulunmuřtur. Genellenebilirlik analizi sonucuna gre en byk varyans kaynaęı sunum olarak belirtilmiřtir. ok yzeyli Rasch analizi sonucu puanlamaların birbiriyle tutarlı olduęu, puanlamaların sunu ve maddelere gre anlamlı bir farklılık gsterdięi ancak uygulamaya gre anlamlı bir farklılık gstermedięi sonucuna ulařılmıřtır. Gz izleme metriklerinden ortalama sabitletme sreleri ok yzeyli Rasch analiziyle incelenmiř ve puanlayıcıların ortalama sabitletme srelerine gre anlamlı bir farklılık gstermedięi bulunmuřtur. Sonu olarak, genellenebilirlik ve ok yzeyli Rasch sonularının birbiriyle uyumlu olduęu gzlemlenmiřtir. Bu doęrultuda gz izleme metrislerinin, puanlayıcılar arası uyumu arařtırmak iin alternatif bir yntem olarak kullanılabileceęi sylenbilir.

Anahtar szckler: gz izleme teknięi, puanlayıcılar arası uyum, genellenebilirlik kuramı, ok yzeyli rasch modeli, puanlayıcılar arası gvenirlik, performans deęerlendirme

Abstract

In this study, the fit between raters, coefficients of fit, generalizability and many facet Rasch were investigated in the performance evaluation process. For this purpose, to evaluate the presentation performance of the students, the relationships between the scores given by the same raters in the classroom during the performance and the scores given on the screen by eye tracking technique were examined. The scores obtained as a result of the in-class evaluation and the evaluation on the screen with the eye tracking technique were analyzed according to the inter-rater fit statistics in the classical test theory, the generalizability theory and the many facet Rasch model. It was found that there was a good agreement between the raters in the presentations of idealism and realism. According to the results of the generalizability analysis, the biggest source of variance was the presentation. As a result of the many facet Rasch model, it was found that the scores were consistent with each other and the scores differed significantly according to the presentation and the items, but did not differ significantly according to the application. The mean fixation times of eye tracking metrics were examined by many facet Rasch model analysis and it was found that the raters did not show a significant difference according to the mean fixation times. As a result, it was found that generalizability and multi-surface Rasch model results were compatible with each other. The study indicated that it could be used as an alternative method to investigate the compatibility between eye tracking metrics and raters.

Keywords: eye tracking, inter-rater agreement, generalizability theory, many facet rasch model, inter-rater reliability, performance assessment

Teşekkür

Tezimin her aşamasında beni destekleyen, tezimi tamamlamam sürecinde verdiği dönütler ve gösterdiği hoşgörü ile beni motive eden, birçok zaman kıymetli vaktini ayırarak karşılaştığım problemlere çözüm yolları oluşturarak devam etmemi sağlayan bir danışmanım olması şansına sahip oldum. Çok Değerli Sayın Danışmanım Prof.Dr. Nuri DOĞAN hocama çok teşekkür ederim.

Doktora eğitimim boyunca aldığım derslerde, doktora yeterlilik sınavı jürimde, tez izleme sürecimde ve tez savunma jürimde yer alan Kıymetli Hocam Prof.Dr. Hülya KELECİOĞLU'na akademik donanımda katkıları, tezime verdiği dönütler ve her zaman gösterdikleri desteklerinden dolayı çok teşekkür ederim. Hacettepe Üniversitesi'nde hem bir öğrenci hem bir araştırma görevlisi olarak yer aldığım süreç boyunca akademik donanımda katkıda bulunan aynı zamanda babacan tavrını her daim hissettiren Sayın Prof. Dr. Selahattin GELBAL hocama çok teşekkür ederim. Doktora yeterlilik sınavı jürimde yer alan, tez izleme ve savunma sürecimde verdiği dönütlerle beni destekleyen, kıymetli vaktini ayıran Sayın Prof.Dr. İsmail KARAKAYA Hocama çok teşekkür ederim. Kıymetli vaktini ayırarak tez savunma jürimde yer alan ve değerli katkılarıyla tezimin zenginleşmesini sağlayan Sayın Prof. Dr. Hakan Yavuz ATAR hocama çok teşekkür ederim.

Bu süreçte her aşamada yanımda olan, her sorunumu çözen, analizlerimi yaparken çok kıymetli vaktini ayırarak yardım eden, hem de bir abla gibi sevgisiyle sarıp sarmalayan Sayın Dr. Öğr. Üyesi Sümeyra SOYSAL hocama çok teşekkür ederim. Hacettepe Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü ve Eğitim Bilimleri Enstitüsü Değerli hocaları ve idari personellerine çok teşekkür ederim. Tezimi dil, anlatım ve düzen açısından inceleyerek çok değerli katkılarda bulunan aynı zamanda devamlı beni motive ederek devam etmemi sağlayan Canım arkadaşım Dr. Öğr. Üyesi Elif Özlem ARDIÇ KEMERKAYA' ya ve Doç.Dr. Meltem YURTÇU' ya çok teşekkür ederim.

Beni türlü zorluklar ve emeklerle büyütüp, okutan ve her konuda desteğini arkamda hissettiğim canım aileme anneme, babama, kardeşlerime çok teşekkür ederim. Bu zorlu süreçte tezimin tamamlanmasında büyük desteği olan ve benimle birlikte çalışarak çalışma motivasyonumu sağlayan eşim Faruk DEMİRBAŞ'a çok teşekkür ederim. Canım kızıma yaşama sevincim ve mutluluğum olduğu için çok teşekkür ederim.

Doktora eğitim sürecim boyunca desteklerinden dolayı TÜBİTAK 'a teşekkür ederim.

İçindekiler

Kabul ve Onay.....	ii
Öz.....	iii
Abstract.....	iv
Teşekkür.....	v
Şekiller Dizini.....	x
Simgeler ve Kısaltmalar Dizini.....	xi
Bölüm 1 Giriş.....	1
Problem Durumu.....	2
Araştırmanın Amacı ve Önemi.....	5
Araştırma Problemi.....	5
Alt Problemler.....	5
Sayıtlılar.....	6
Sınırlılıklar.....	6
Tanımlar.....	6
Bölüm 2 Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar.....	7
Puanlayıcılar Arasındaki Uyum.....	7
Puanlayıcılar Arası Uyumu Araştırmak İçin Kullanılan Teknikler.....	8
Genellenebilirlik Kuramı.....	11
Çok Yüzeyle Rasch Modeli.....	12
Göz İzleme.....	14
İlgili Araştırmalar.....	15
Bölüm 3 Yöntem.....	26
Araştırmanın Türü.....	26
Çalışma Grubu.....	26
Veri Toplama Süreci.....	27

Veri Toplama Araçları	28
Göz İzleme Yöntemi.....	29
Verilerin Analizi	33
Bölüm 4 Bulgular ve Tartışma	35
Birinci Alt Probleme İlişkin Bulgular.....	37
İkinci Alt probleme İlişkin Bulgular.....	39
Üçüncü Alt Probleme İlişkin Bulgular	41
Dördüncü Alt Probleme İlişkin Bulgular	49
Beşinci Alt Probleme İlişkin Bulgular	54
Tartışma.....	65
Bölüm 5 Sonuç ve Öneriler.....	70
Araştırmaya Dönük Öneriler.....	72
Uygulamaya Dönük Öneriler	74
EK-A: Sunum Değerlendirme Dereceli Puanlama Anahtarı	lxxxvii
EK-B: Etik Komisyonu Onay Bildirimi	lxxxviii
EK-C: Etik Beyanı.....	lxxxix
EK-Ç: Yüksek Lisans/Doktora Tez Çalışması Orijinallik Raporu	xc
EK-D: Thesis/Dissertation Originality Report.....	xcı
EK-E: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı	xcii

Tablolar Dizini

Tablo 1 <i>Kappa İstatistiği Değer Aralıkları</i>	9
Tablo 2 <i>Sınıf İçi Korelasyon İstatistiği Değer Aralıkları</i>	10
Tablo 3 <i>Uzman Grubunun Özellikleri</i>	28
Tablo 4 <i>Sınıf İçi Uygulama ve Göz İzleme Uygulaması Sonucunda Üç Puanlayıcıdan Elde Edilen Puanların Betimsel İstatistikleri</i>	36
Tablo 5 <i>Cohen Kappa Analizi Sonuçları</i>	37
Tablo 6 <i>Üç Puanlayıcı İçin Fleiss Kappa Analiz Sonuçları</i>	38
Tablo 7 <i>Sınıf İçi Korelasyon Analizi Sonuçları</i>	38
Tablo 8 <i>Tamamen Çaprazlanmış (UxPxSxM) Desene İlişkin G Çalışması Sonuçları</i>	40
Tablo 9 <i>Sınıf İçi ve Göz İzleme Uygulaması Puanlarının Çok Yüzeyle Rasch Analiziyle Elde Edilen Puanlayıcı Yüzeyi Ölçüm Parametreleri</i>	44
Tablo 10 <i>Sınıf İçi ve Göz İzleme Uygulaması Puanlarının Çok Yüzeyle Rasch Analiziyle Elde Edilen Uygulama Yüzeyi Ölçüm Parametreleri</i>	46
Tablo 11 <i>Sınıf İçi ve Göz İzleme Uygulaması Puanlarının Çok Yüzeyle Rasch Analiziyle Elde Edilen Sunu Yüzeyi Ölçüm Parametreleri</i>	47
Tablo 12 <i>Sınıf İçi ve Göz İzleme Uygulaması Puanlarının Çok Yüzeyle Rasch Analiziyle Elde Edilen Madde Yüzeyi Ölçüm Parametreleri</i>	48
Tablo 13 <i>Puanlama Sonuçlarının Çok Yüzeyle Rasch Modeline Göre Kategori Analiz Sonucu İstatistikleri</i>	49
Tablo 14 <i>Ortalama Sabitlenme Süreleri Metrikleri Çok Yüzeyle Rasch Analiziyle Elde Edilen Puanlayıcı Yüzeyi Ölçüm Parametreleri</i>	51
Tablo 15 <i>Ortalama sabitlenme Süreleri Metrikleri Çok Yüzeyle Rasch Analiziyle Elde Edilen Odak Noktası Yüzeyi Ölçüm Parametreleri</i>	52
Tablo 16 <i>Ortalama sabitlenme Süreleri Metrikleri Çok Yüzeyle Rasch Analiziyle Elde Edilen Sunu Yüzeyi Ölçüm Parametreleri</i>	53
Tablo 17 <i>Göz İzleme Uygulaması Ortalama sabitlenme Süreleri Metrikleri Çok Yüzeyle Rasch Analiziyle Elde Edilen Kategori İstatistikleri</i>	54
Tablo 18 <i>İdealizm Sunumunda Toplam Sabitlenme Sürelerine(Sn) Ait Betimsel İstatistikler</i>	55
Tablo 19 <i>Yaratıcılık Sunumu Toplam Sabitlenme Sürelerine Ait Betimsel İstatistikler</i>	58

Tablo 20 <i>Realizm Sunumu Toplam Sabitleme Sürelerine Ait Betimsel İstatistikler</i>	61
Tablo 21 <i>Göz İzleme Uygulaması ile Elde Edilen Puanlayıcılara Ait Metrikler Genel Tablosu</i>	64

Şekiller Dizini

Şekil 1 <i>İdealizm Sunumu İlgili Alanları</i>	30
Şekil 2 <i>Yaratıcılık Sunumu İlgili Alanları</i>	31
Şekil 3 <i>Realizm Sunumu İlgili Alanları</i>	32
Şekil 4 <i>Çok Yüzeyle Rasch Analizi Değişken Haritası</i>	42
Şekil 5 <i>Göz İzleme Tekniğiyle Elde Edilen Ortalama Sabitletme Metriklerinin Çok Yüzeyle Rasch Analizi Değişken Haritası</i>	50
Şekil 6 <i>İdealizm Sunumu Yol Haritası</i>	56
Şekil 7 <i>İdealizm Sunumu Sıcaklık Haritası</i>	57
Şekil 8 <i>Yaratıcılık Sunumu Yol Haritası</i>	59
Şekil 9 <i>Yaratıcılık Sunumu Sıcaklık Haritası</i>	60
Şekil 10 <i>Realizm Sunumu Yol Haritası</i>	62
Şekil 11 <i>Realizm Sunumu Sıcaklık Haritası</i>	63

Simgeler ve Kısaltmalar Dizini

İLÖ1: İlgı Alanı Öğrenci İdealizm Sunumu

İİÖ2: İlgı Alanı Öğrenci Yaratıcılık Sunumu

İİÖ3: İlgı Alanı Öğrenci Realizm Sunumu

İLt1: İlgı Alanı Tahta İdealizm Sunumu

İLt2: İlgı Alanı Tahta Yaratıcılık Sunumu

İLt3: İlgı Alanı Tahta Realizm Sunumu

KTK: Klasik Test Kuramı

MTK: Madde Tepki Kuramı

SİD: Sınıf İçi Deęerlendirme

P1: 1.puanlayıcı

P2: 2.puanlayıcı

P3: 3.puanlayıcı

Bölüm 1

Giriş

Eğitimde; öğrenciyi merkeze alan, üst düzey becerileri ölçebilen, öğrenmede ürün ve süreci birlikte değerlendirebilen yaklaşımlara olan ihtiyaç performans değerlendirmeyi gündeme getirmektedir. Performans değerlendirme, üst düzey becerilerin, belirli kriter ve ölçme araçlarına dayalı olarak ölçülme süreci olarak tanımlanabilir (Kutlu ve ark., 2009). Performansın puanlanmasında, genellikle, kontrol listesi, davranış sıklığı ölçeği ve dereceli puanlama anahtarı (rubrik) gibi puanlama araçları kullanılmaktadır (Kutlu ve ark., 2009). Dereceli puanlama anahtarlarında, performans görevine ilişkin ölçütler ve performansın bu ölçütlere ilişkin nitelik derecesi tanımlanmaktadır (Goodrich, 1997). Ölçülecek performansın yapısı ve değerlendirmenin amacına bağlı olarak analitik ve bütünsel (holistik) olmak üzere iki çeşit dereceli puanlama anahtarı kullanılmaktadır (Doğan, 2020; Mertler, 2000; Popham, 1997). Analitik puanlama anahtarı, ölçülen özellik öğelerine ayrıştırılabildiği ve daha ayrıntılı puanlama yapılmak istenildiğinde kullanılmaktadır (Haladyna, 1997). Holistik puanlama anahtarında ise performans farklı boyutlara ayrılmadan belirlenir ve tek bir puan verilir (Kutlu ve ark., 2009). Performans değerlendirmeden elde edilen sonuçların güvenilirliği ve geçerliğinin artırılması için dereceli puanlama anahtarlarının kullanılması önerilmektedir (Dunbar ve ark, 2006). Ancak performansın puanlanmasında sınırlı sayıda, sadece işlevsel olan ve gözlenebilir ölçütlerin tanımlanması, ölçütlerin hangi düzeydeki puana denk geleceğinin belirlenmesi, düzeyler arasındaki uzaklığın eşit tutulması gibi süreçlerin belirlenmesi oldukça zor ve zaman alıcı bir süreçtir. Ayrıca performans ölçüt ve düzeylerinin daha açık ve net bir şekilde belirtildiği rubrikler ile yapılan puanlamalarda bile öznellik (sübjektiflik) söz konusu olabilmektedir. Bu nedenle puanlayıcı güvenilirliği, performans değerlendirmenin yapıldığı tüm araştırmalarda önemli bir faktördür. Bir başka ifadeyle, performans değerlendirmede puanlayıcılar önemli bir hata kaynağı olabilir (Brennan ve ark., 1995; Bıkmaz Bilgen ve Doğan, 2017; Congdon & MeQueen, 2000). Puanlamanın objektifliği ise ölçme sonuçlarının güvenilirliğini doğrudan etkilemektedir (Baykul, 1999).

Performans değerlendirme ile ilgili en önemli kaygı, performansın objektif puanlanmasının zor olmasından kaynaklanmaktadır (Romagnano, 2001). Performans değerlendirilirken, farklı puanlayıcıların aynı cevaba aynı puanı vermesi oldukça güçtür (Turgut ve Baykul, 2014). Puanlayıcı etkisi, ölçme sonuçlarına karışan, istenmedik hata kaynaklarıdır. Bir başka ifadeyle, puanlayıcılar bireylerin performans puanlarını etkileyen temel faktörlerdendir (Farrokhi ve ark., 2011). Hata kaynağının puanlayıcılar olduğu durumlar için puanlayıcı güvenilirliğinin incelenmesi önerilmektedir (Cohen ve ark., 1996). Bu nedenle, farklı puanlayıcıların bireylerin performansına ilişkin yaptıkları puanlamalar arasındaki tutarlık (güvenirlilik) önem arz etmektedir. Puanlama sürecinde ölçme sonuçlarına puanlayıcıdan kaynaklı hataların karışması ise elde edilen sonuçların güvenirliliği ve geçerliliği için önemli bir tehdit oluşturmaktadır. Bu nedenle, puanlayıcıdan kaynaklı hataların ölçümlere karışma düzeylerinin azaltılması veya kontrol altına alınması gerekmektedir (Linacre, 2006). Puanlayıcı etkileri; puanlayıcı katılımı ve cömertliği, merkeze yönelme etkisi, halo etkisi, ranj sınırlaması, tutarsızlık ve yanlılık şeklinde gözlenebilir (Engelhard ve ark., 2018).

Problem Durumu

Puanlayıcı güvenirliliği, puanlayıcı içi güvenirlilik ve puanlayıcılar arası güvenirlilik olarak ifade edilmektedir (Bıkmaz Bilgen ve Doğan, 2017). Puanlayıcı içi güvenirlilik, bir puanlayıcının aynı bireylere farklı zamanlarda verdiği puanlar arası korelasyondur. Puanlayıcılar arası güvenirlilik ise iki ya da daha fazla puanlayıcının aynı bireylere ilişkin birbirinden bağımsız olarak yaptığı puanlamalar arası tutarlılığın derecesi olarak ifade edilmektedir (Aiken, 2000; Anastasi & Urbina, 1997; Cohen ve ark., 1996; Kutlu ve ark., 2009; Viera & Garret, 2005). Performans değerlendirme sürecinde ölçme sonuçlarına karışan puanlayıcı etkisinin azaltılması için birden fazla puanlayıcının kullanılması (Atılğan, 2004; 2008; Bıkmaz Bilgen ve Doğan, 2017; Deliceoğlu, 2009; Ebel, 1951; Johnson ve ark., 2000; Kubiszyn & Borich, 2016; Taşdelen, 2009; Turgut, 1993) ve dereceli puanlama anahtarlarının kullanılması (Kutlu ve ark., 2009) önerilmektedir.

Performans deęerlendirmenin bilişsel, duyuşsal ve psikomotor alan becerilerini birlikte ölçebilmesi, bireyleri üst düzey düşünmeye ve gerçek yaşam sorunları ile ilgilenmeye yönlendirmesi üstün yönleridir. Ancak puanlamanın objektif ve kolay olmaması önemli dezavantajdır. Daha önce de dile getirildięi gibi dereceleme ölçęęi, puan cetveli veya dereceli puanlama anahtarı kullanılsa dahi istenilen düzeyde objektiflik sağlanamamaktadır. Bu nedenle, güvenilir puanlama yapmayı kolaylaştıracak yeni yaklaşımlara ve yöntemlere gereksinim duyulmaktadır. Bu ihtiyacın da bir yansıması olarak son yıllarda özellikle hem performans deęerlendirmede hem de puanlama güvenilirliğini belirleme sürecinde teknolojinin bilimsel çalışmalara nasıl entegre edilebileceęi araştırılmaktadır (Erdoęmuş ve Çaęıltay,2009).

Teknoloji; bilimin farklı alanlardaki sorunlarını göz önüne alarak, bilim ve uygulama arasında köprü işlevi gören makineler, işlemler, yöntemler, süreçler, sistemler, yönetim ve denetim mekanizmalarının bileşeni olarak tanımlanabilmektedir (Alkan, 2005). Eęitimle ilgili bilimsel çalışmalarda yararlanılan teknolojiler, eęitim teknolojileri olarak ifade edilmektedir (Anglin,1995). Eęitim teknolojisi kavramı ise üretilen teknolojilerin, öğrenme–öğretme süreçlerinin tasarlanması, uygulanması ve geliştirilmesi sürecine dâhil edilmesi olarak tanımlanmaktadır (Alkan, 2005; Erdoęmuş ve Çaęıltay, 2009). Eęitim teknolojileri; öğrenme yaklaşımları, performans teknolojileri ve öğretim sistemlerinden oluşmaktadır (Çakır ve ark., 2013; Töngel ve ark., 2020). Farklı bilişsel ve duyuşsal özelliklerin ölçüldüğü elektronik sınavlar, otomatik puanlama sistemleri, bireysel kişilik testleri, üst düzey becerilerin ölçüldüğü elektronik portfolyo yazılımları, ölçme ve deęerlendirme sürecinde kullanılan eęitim teknolojileri arasında yer almaktadır (Ballard, 2017; Morgan & O'Reilly, 2020). Eęitim alanındaki çalışmalar incelendiğinde, kullanılan yeni yaklaşımlardan birisinin Göz İzleme Teknięi olduęu görülmektedir.

Göz İzleme Teknięi, 1980'li yıllardan itibaren eęitim araştırmalarında kullanılmaya başlamıştır (Durna ve Arı, 2016). Ancak yapılan çalışmalar son yıllarda artış göstermiştir. Göz İzleme Yöntemi, belirlenen çerçeve içerisinde yer alan metin, video veya fotoęrafta

bireyin nerelere ve ne kadar süre ile baktığı, dikkatini hangi noktalara yoğunlaştırdığı hakkında bilgi vermektedir (Alemdağ ve Çağıltay, 2018; Wotschack, 2009). Bireylere ilişkin göz izleme sonuçları, dikkatlerini nerelere yoğunlaştırdıkları, hangi öğeleri ilginç buldukları vb. konusunda çıkarım yapmayı mümkün kılmaktadır (Duchowski, 2017). Bir başka ifadeyle, kullanıcıların nereye baktığının gözlenebilmesi ile ne düşündükleri bilinemese de bununla ilgili fikir yürütülebilir (Peignot & Anderson, 1999). Bu bağlamda, performans değerlendirme sürecinde göz izleme yönteminin uygulanabilirliğinin ve puanlayıcıların sınıf içi gözlemlere ve göz izleme yöntemine göre verdikleri puanlar arasındaki uyumun incelenmesinin önemli olduğu düşünülmektedir.

Son yıllarda, hem ülkemizde hem de uluslararası düzeyde eğitim alanında yapılan tezlerde, Göz İzleme Tekniğinin eğitim araştırmalarına katkısı gözlemlenmektedir (Armut, 2021; Ballard, 2017; Coşkun, 2019; Coşkuner, 2022; Nalçacıer, 2014; Sağlam, 2022; Utku, 2019; Turan, 2018). Yapılan çalışmalar incelendiğinde, performans değerlendirme sürecinde puanlayıcıların göz hareketleri ölçüm verileri, farklı parametrelere göre incelendiği (Wang ve ark., 2017; Ballard, 2017; Ma & Winke 2022) göz izleme metriklerinin puanlayıcıların vermiş oldukları puanlarla uyumlu olduğu ve puanlayıcı verileri yerine de kullanılabilceği belirtilmektedir (Stuart ve ark., 2018). Ancak, göz izleme tekniği ile elde edilen sonuçlarının güvenilirliğinin belirlenmesine ve puanlayıcılar arası uyuma katkı sağlayıp sağlamadığına yönelik az sayıda araştırmanın yapıldığı ve bu alanda yapılacak yeni araştırmalara ihtiyaç olduğu belirtilmiştir (Friedman ve ark., 2022). Göz izleme yönteminin performans değerlendirme sürecinde kullanımıyla ilgili literatür incelendiğinde, sınıf içi öğrenci performansını ve sunum becerisini ölçen bir çalışmaya rastlanmamıştır. Göz izleme yönteminden elde edilen göz izleme ölçümlerinin, belirli bir amaç için belirlenen ölçütlere dayalı olarak, bireylerin performanslarına puanlayıcıların vermiş oldukları puanların, geçerlik ve güvenilirlik açısından incelenmesine ne tür bir etkisinin olacağı, araştırılmaya devam edilen konulardan biridir.

Araştırmanın Amacı ve Önemi

Araştırmanın genel amacı teknolojinin imkânlarından yararlanarak performans değerlendirme sürecinde daha güvenilir sonuçların nasıl elde edilebileceğine ilişkin güncel ve özgün bir çalışma yapmaktır. Performans değerlendirme sürecinde, göz izleme tekniğini, puanlayıcılar arası uyum için kanıt toplama aracı olabileceği varsayılmıştır. Bu doğrultuda çalışmanın temel amacı, göz izleme yöntemiyle elde edilen veriler ile puanlayıcıların sunum değerlendirme dereceli puanlama anahtarına göre yaptığı puanlamalar arasındaki uyumu incelemektir. Dolayısıyla bu çalışmada, “Puanlayıcı uyumu çalışmalarında kullanılan belirli istatistiksel yöntemlerle elde edilen puanlayıcılar arası uyum sonuçlarının göz izleme tekniğinden elde edilen sonuçlarla desteklenmesi mümkün müdür?” sorusuna cevap aranmaktadır. Bu araştırma sonucunda performans değerlendirme sürecinde göz izleme tekniği verilerinin kullanılmasının puanlayıcılar arası uyum için destekleyici kanıt oluşturması ve puanlama güvenilirliğini değerlendirmede yararlı olması beklenmektedir.

Bu çerçeveye bağlı olarak, araştırmanın önemi performans değerlendirme sürecinde göz izleme tekniğinden yararlanılarak, puanlayıcılar arası uyum hakkında kanıt toplayarak performans değerlendirmede verilen puanların güvenilirliğini destekleyen sonuçlar elde etmek olarak ifade edilebilir.

Araştırma Problemi

Öğrencilerin sunum yapma performansını değerlendirme sürecinde, aynı puanlayıcıların performans sergileme esnasında sınıf içinde verdikleri puanlarla göz izleme tekniği ile ekran üzerinden verdikleri puanlar arasındaki ilişkiler nasıldır?

Alt Problemler

Araştırmanın ana problemine bağlı olarak oluşturulan alt problemler aşağıda sıralanmıştır:

1. Performans deęerlendirme s¼recinde Cohen Kappa, Fleiss Kappa ve Sınıf İçi Korelasyon ile elde edilen puanlayıcılar arası uyum d¼zeyleri nasıldır?
2. Performans deęerlendirme s¼recinde elde edilen puanların, genellenebilirlik kuramı uygulamasıyla elde edilen varyans kaynaklarının deęeri nasıldır?
3. Performans deęerlendirme s¼recinde elde edilen puanlara uygulanan, çok y¼zeyli Rasch modeli sonuçları nasıldır?
4. G¼z izleme sonucunda puanlayıcılardan elde edilen g¼z izleme metriklerine uygulanan, çok y¼zeyli Rasch modeli sonuçları nasıldır?
5. G¼z izleme sonucunda puanlayıcılardan elde edilen g¼z izleme metrikleri nasıldır?

Sayıtlılar

Performans deęerlendirme s¼recine katılan puanlayıcıların G¼z izleme Y¼ntemiyle elde edilen verileri etkileyecek bir g¼z kusuruna sahip olmadığı varsayılmaktadır.

Sınırlılıklar

- 1.Çalışma, ¼ç farklı performans görevi ile sınırlıdır.
- 2.Çalışma, ¼lçme ve deęerlendirme alanında uzman ¼ç puanlayıcı ile y¼r¼t¼lm¼şt¼r.
3. Çok y¼zeyli Rasch analizi varsayımlarının karşılandığı kabul edilmiştir.
4. Giyilebilir g¼z izleme cihazı yerine taşınabilir g¼z izleme cihazı kullanılmıştır.

Tanımlar

G¼z izleme metrikleri: g¼z izleme sonucunda elde edilen istatistiksel verilerdir (Alemdaę ve Çaęıltay, 2018; Lai ve ark., 2013).

Bölüm 2

Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar

Puanlayıcılar Arasındaki Uyum

Puanlayıcıların tutarlı puanlar vermesi beklenen bir durumdur. Bu koşulun sağlanması için araştırmalarda birden fazla puanlayıcının sürece katılması, puanlayıcıların puanlama eğitimi alması, otomatik puanlama yazılımlarının kullanılması gibi farklı yollar izlenmektedir (Friedman ve ark., 2022; İlhan, 2015; Şata, 2019).

Puanlayıcıların ölçülmek istenen değişkeni tam ve doğru ölçmesi, ölçme sürecine hataların karışmaması, puanlayıcı güvenilirliği kavramını ifade etmektedir. Dereceli puanlama anahtarıyla yapılan performans değerlendirme süreçlerinde, ölçme sonuçlarının güvenilirliğini etkileyen en önemli hata kaynağı puanlayıcılardır (Myford & Wolfe, 2004). Literatürde puanlayıcıların dikkatsizlik, gözlem yetersizliği, yanlılık, halo etkisi, merkeze kayma etkisi, genelleme hatası gibi olumsuz faktörlere sahip olması, ölçme sonuçlarının güvenilirliğini olumsuz etkilediği belirtilmektedir (Engelhard ve ark., 2018; Turgut & Baykul, 2014). Performans değerlendirme sürecinin önemli bir faktörü olan puanlayıcıların özellikleri yapılan çalışmalarda araştırılmaktadır. Puanlayıcıların hazırbulunuşluk düzeyleri, bireysel özellikleri ve puanlama becerilerinin puanlama sonuçlarını farklılaştırdığı yorumu yapılabilmektedir (Barkaoui, 2008; Engelhard ve ark., 2018; Friedman ve ark., 2022).

Puanlayıcı katılımı ve puanlayıcı cömertliği puanlayıcıların, puanlama ölçütüne göre yaptığı puanlamalarda diğer puanlayıcılardan daha düşük veya daha yüksek puan verme eğilimi olarak yorumlanmaktadır (Engelhard ve ark.,2018). Bu durumun puanlama güvenilirliğini olumsuz etkilediği belirtilmiştir. Halo etkisi; puanlayıcının, performans puanlama ölçütüne farklı bakış açısıyla yaklaşması nedeniyle, farklı nedenlere bağlı puan vermesidir. Merkeze eğilim davranışı; puanlama ölçütünün kategorilerinin ortadaki değerinin daha fazla işaretlenmesidir. Yanlılık; puanlayıcının, puanlanan bireylerin yaş, cinsiyet veya etnik köken gibi özelliklerine göre farklı puan verme durumu olarak tanımlanabilmektedir. Puanlayıcı

tutarsızlığı; puanlayıcının rastgele puan vermesi, bazen puanlama ölçeğine diğer puanlayıcılardan yüksek bazen ise düşük puan vermesi durumudur (Myford & Wolfe, 2004; Kumar, 2005; Wiseman, 2008). Performans değerlendirme sürecinde puanlayıcı güvenilirliğini araştırmak için çeşitli istatistikler kullanılmaktadır. KTK ya göre, MTK ya göre ve Genellenebilirlik kuramına göre farklı yöntemler geliştirilmiştir (Cronbach, 1972; Johnson ve ark., 2008; Ma & Winke, 2022; Winke & Lim, 2015). Aşağıda bu araştırmada kullanılan puanlayıcı güvenilirliği belirlemek teknikleri hakkında gerekli bilgiler verilmiştir.

Puanlayıcılar Arası Uyumu Araştırmak İçin Kullanılan Teknikler

Puanlayıcılar arası güvenilirlik, şansla uyum kontrol altına alındığında, hedef puanlayıcı evreninden seçilen herhangi iki puanlayıcının, hedef özne evreninden seçilen herhangi bir özneyi aynı öncelik seviyesine atama eğilimi olarak tanımlanmaktadır (Karakaya, 2020). Puanlayıcılar arası güvenilirlik, puanlayıcıların verdikleri puanların birbirleriyle tutarlı olması ile ilişkilendirilmektedir. Bu durum Puanlayıcılar arası uyum kavramını oluşturmaktadır. Puanlayıcılar arası uyum farklı istatistiksel yöntemlerle araştırılmaktadır. Cohen (1960) tarafından iki puanlayıcı arasındaki uyumun belirlenmesi için Kappa İstatistiği geliştirilmiştir. Kappa istatistiği κ aşağıda yer alan formülle hesaplanmaktadır (Landis & Koch, 1977).

$$\kappa = \frac{P - P_E}{1 - P_E}$$

P =Gözlenen uyumluluk oranı, P_E = Tesadüfi uyumluluk oranı

Kappa istatistiği tesadüfi hatanın karışması sonucunda beklenenden daha yüksek bir değere sahiptir. Değerlendirmenin bağımsız olmaması ve yanlılık kappa katsayısının değerini etkilemektedir. Fleiss Kappa ikiden fazla puanlayıcı arasındaki uyumu ölçmek için, şans başarısı faktörü göz önüne alınarak geliştirilmiştir (Fleiss, 1971; Fleiss ve ark., 2003).

Kappa analizi sonucu elde edilen verilerin yorumlanması için kullanılan sınır değerler

Tablo 1 'de verilmiştir (Landis & Koch,1977).

Tablo 1

Kappa İstatistiği Değer Aralıkları

κ	Uyumun Yorumu
0,00 – 0,20	Önemsiz
0,21 – 0,40	Düşük
0,41 – 0,60	Orta
0,61 – 0,80	Önemli
0,81 – 1,00	Çok Yüksek

Tablo 1'de Kappa istatistiği ile elde edilen katsayının nasıl yorumlanacağı açıklanmaktadır. Performans değerlendirme sürecinde elde edilen puanlar kappa analizi incelendiğinde, $\kappa = 0,00-0,20$ önemsiz bir uyum olduğunu, $\kappa = 0,21-0,40$ orta düzeyde bir uyum olduğunu, $\kappa = 0,41-0,60$ orta düzeyde bir uyum olduğunu, $\kappa = 0,61-0,80$ elde edilen kappa katsayısının önemli bir uyum olduğunu göstermektedir. Bunun yanında $\kappa = 0,81-1,00$ olduğunda ise puanlayıcılar arasında çok yüksek bir uyum olduğu şeklinde yorumlanmaktadır (Landis & Koch,1977).

Cohen (1968), sıralama ölçeğindeki puanlar arasındaki uyumu ağırlıklandırılmış kappa ile belirlemeyi önermektedir. Puanlayıcılar arasındaki uyumun incelenmesinde, puanlayıcı sayısı ve ölçek türlerine göre farklı uyum istatistikleri kullanılmaktadır. Ağırlıklandırılmamış kappa bütün uyumsuzlukları eşit olarak ele alırken, ağırlıklandırılmış kappa uyumsuzlukları önemine göre değerlendirir. Bu nedenle, ağırlıklandırılmamış kappa, sıralama düzeyindeki ölçekler için uygun değildir. Puanlayıcılar arası güvenirliliğin belirlenmesinde kullanılan bir diğer yöntem de sınıf-içi korelasyon katsayısıdır (Barkto,1966). İki ve ikiden fazla puanlayıcının toplam puanlar üzerinden en az sıralı kategorik veriler için puanlayıcılar arası uyumu inceler. Sınıf içi korelasyon katsayısı ile aynı birey farklı puanlayıcı ve aynı birey farklı zamanda aynı puanlayıcı puanları arasındaki uyum incelenmektedir

(Bilgen ve Dođan, 2017). Deđişkenlik kaynađına göre farklı modeller oluşturulmaktadır. Sınıf ii korelasyon katsayısı genel formülünde puanlayıcılar arası varyansın, puanlayıcılar ii varyansa oranı olarak ifade edilmektedir (Barkto, 1966). Sınıf ii korelasyon katsayısı ařađıdaki formülle ifade edilmektedir.

$$\rho = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_W^2}$$

Kullanım amacına uygun modelin seilmesi, yorumların dođru bir şekilde elde edilmesi iin önem arz etmektedir (Karakaya, 2020). Likert tipi ölçeklerin ve sıralı kategorik verilerin güvenilirlik analizlerinde sınıf ii korelasyon katsayısı kullanılabilir (Kutlu ve ark., 2009). Sınıf ii korelasyon katsayısı 0 ile 1 arasında deđerler alır. Sınıf ii korelasyon katsayısının temeli, varyans analizine ve farklı varyans bileşenlerinin tahmin edilmesine dayanmaktadır. Deneysel desenin özelliđine göre farklı varyans modelleri tanımlanmaktadır (Barkto, 1966). Sınıf ii korelasyon analizi sonucu elde edilen verilerin yorumlanması iin kullanılan sınır deđerler Tablo 2 'de verilmiştir (Fleiss ve ark., 2003).

Tablo 2
Sınıf İi Korelasyon İstatistiđi Deđer Aralıkları

r	Uyumun yorumu
r < 0,40	Zayıf uyum
r=0,41- 0,59	Orta uyum
r= 0,60- 0,74	İyi uyum
r> 0,75	Mükemmel uyum

Tablo 2'de sınıf ii korelasyon katsayısının (ICC) sınır deđerleri yer almaktadır. Performans deđerlendirme sürecinde puanlayıcılar arasındaki uyum sınıf ii korelasyon katsayısı ile incelendiđinde, r<0,40 olduđunda zayıf bir uyum olduđu, r= 0,41-0,59 arasında orta derecede bir uyum olduđu, r=0,60-0,74 arasında iyi uyum olduđunu ve r> 0,75

olduğunda puanlayıcılar arasında mükemmel uyum olduğu şeklinde yorumlanmaktadır (Fleiss ve ark., 2003).

Genellenebilirlik Kuramı

Varyans analizine dayalı olarak geliştirilen Genellenebilirlik Kuramı (Brennan, 1992; Shavelson & Webb, 1991), Genellenebilirlik ve Karar çalışmalarıyla güvenilir gözlemlerin elde edilmesi, elde edilen puanların güvenilirliğinin yorumlanması, tesadüfi ve sistematik hata kaynaklarının belirlenmesinde etkili olduğu belirtilmiştir. Shavelson ve Webb (1991), KTK ile Genellenebilirlik Kuramını karşılaştırdıklarında; Genellenebilirlik Kuramını 'nın varyans bileşenlerini tek bir analizle ele alabildiğini, her bir varyans kaynağının büyüklüğünün belirlenebildiğini, birey performanslarına dayalı bağıl kararlar (G katsayısı) ve mutlak kararlar (phi katsayısı) ile ilgili güvenilirlik katsayılarının hesaplanmasını sağladığını ifade etmiştir. Bunun yanında genellenebilirlik kuramının karar çalışmasıyla ölçme hatalarını en aza indirebilecek ölçme düzenlemesini sağladığını belirtmişlerdir. Bu nedenle, Genellenebilirlik Kuramı ile KTK'ya göre daha kapsamlı bir güvenilirlik kestirimi gerçekleştirmek söz konusu olmaktadır. Genellenebilirlik Kuramı, ölçmedeki hata kaynaklarının dayandığı değişkenlik kaynaklarının ortak etkilerini (etkileşimlerini) de dikkate almaktadır. Genellenebilirlik Kuramını hata kaynakları bileşenlerini, test tekrar test güvenilirliği, puanlayıcılar arası güvenilirlik ve iç tutarlık güvenilirliğini birlikte değerlendirebilmektedir. Genellenebilirlik Kuramında gözlemler evreninden alınan örneklemin evrene genellenebilirliği test edilmektedir. Genellenebilirlik Kuramı kapsam geçerliği için de kanıt oluşturduğu iddia edilmektedir (Brennan, 1992; Shavelson & Webb, 1991).

Genellenebilirlik Kuramına göre evrende bulunan birden çok varyans kaynağı, değişkenlik kaynağı (facet=yüzey) olarak adlandırılır. Değişkenlik kaynağı ölçme hatasının olası kaynağı olarak tanımlandığı için değişkenlik kaynağı ile ilgili varyansın küçük olması istenmektedir (Güler, 2008). Maddeye verilen cevaplardaki varyans (değişkenlik) madde

güçlüğünden, bireylerin performansından ve birey ve maddeler arası etkileşimden kaynaklanabilmektedir.

Değişkenlik kaynağının her bir farklı değeri, değişkenlik kaynağının koşulları (conditions) olarak isimlendirilir. Ölçme işleminde birey değişkenlik kaynağının koşulu olarak düşünülebilir. Birey genel olarak ölçme işleminde hedef yani ölçmenin amacı (the object of measurement) olarak belirlenir. Bireyler arasında varyans istenilen bir durum olduğundan, ölçmenin amacı bireyler arasındaki farkı ortaya koymaktır (Güler, 2008).

Genellenebilirlik Kuramında değişkenlik kaynakları sabit ve rastgele olarak tanımlanabilmektedir. Rastgele değişkenlik kaynağı olduğu durumlarda, o değişkenlik kaynağı için tüm durumların yer aldığı evrene genelleme yapabilme imkânı tanımaktadır. Genelleme yapma amacı bulunmayan çalışmalarda ise değişkenlik kaynağı sabit olarak belirtilmektedir (Crocker & Algina, 1986). Genellenebilirlik Kuramında, G çalışması ve karar çalışması iki ana uygulama olarak adlandırılmaktadır. G çalışmasında çeşitli değişkenlik kaynaklarına ait varyans bileşenleri ve ölçme hataları hesaplanır. Karar çalışmasında ise varyans bileşenlerini kullanarak hatayı en aza indirmek için hangi yüzeylerin sayısını nasıl seçmemizin uygun olacağına ilişkin çeşitli sonuçlar elde edilmekte, bu sonuçlara ilişkin G ve Phi değerleri hesaplanmaktadır (Rentz, 1987, akt. Nalbantoğlu, 2009). Phi katsayısı, özellikle, performans ölçümlerinde tercih edilmektedir (Güler, 2008; Lee & Frisbie, 1999).

Çok Yüzeyle Rasch Modeli

İlk Rasch Ölçme Modeli, 1960 yılında George Rasch tarafından, iki kategorili doğru / yanlış şeklinde puanlanan ölçme araçları için geliştirilmiştir. Daha sonra kısmi puanlama modeli (Andrich, 1978), dereceleme ölçeği modeli (Masters, 1982), çok yüzeyle Rasch modeli (Linacre, 2006) geliştirilmiştir. Çok yüzeyle Rasch analizinde, tüm varyans kaynaklarının birbiriyle olan etkileşimi analiz edilmektedir.

Linacre (2018), Rasch modelinin tek boyutlu olarak ele alındığını ifade etmektedir. Ölçülmek istenen yapıya katkı sağlamayan kısımlar artıklar olarak ifade edilip, analize

alınmamaktadır. Rasch artıkları (Rasch residuals), Rasch modeli ile kestirilen raslantısallık (randomness), tek boyutlu Rasch yöntemiyle elde edilen yetenek değişkeni dışında kalan boyutların bileşenleri ve boyut dışı etkenler (rastgele tahmin) gibi faktörler olmak üzere bölümlere ayrılmaktadır (Linacre, 2018). İki yüzeyli Rasch modelin geliştirilmiş hali olan çok yüzeyli Rasch modeli ise ölçüm puanlarının katılımcı ve madde kategorilerinin yanında daha farklı değişkenlik kaynaklarını da incelemektedir. Her bir yüzey, değişkenlik kaynağını ifade etmektedir. İstatistiksel bir model olan çok yüzeyli Rasch, ham puanların doğal logaritmasını alır (log odds). Puanlar dönüştürülerek, logit adı verilen eşit aralıklı ölçek düzeyine getirilirler. Rasch, model veri uyumuna karar vermek için ki-kare uyum istatistiklerinin kullanımını önermektedir. Bunlar iç uyum (infit) ve dış uyum (outfit) olarak ifade edilmektedir (Elhan & Atakurt, 2005). Uyum istatistikleri model veri uyumu hakkında yorum yapma imkânı vermektedir. Her bir yüzeye ait parametre kestirimleri logit cetvel üzerinde yer almaktadır. Görsel olarak yüzeylerin yerleri incelenmekte ve yorumlanmaktadır. Çok yüzeyli Rasch modeli, puanlayıcı kaynaklı farklı faktörler (katılık-cömertlik, halo etkisi, merkeze yönelme ve yanlılık gibi) hakkında bilgi vermektedir (Myford & Wolfe, 2004). Her bir yüzey için 1 ile ∞ arasında değer alan ayırma indeksi ve 0 ile 1 arasında değer alan güvenilirlik katsayısı elde edilmektedir.

Birey yüzeyi için güvenilirlik değerinin yüksek olması, bireylerin birbirinden ayırt edilebildiğine işaret eder. Madde yüzeyi için elde edilen ayırma indeksi ve güvenilirlik katsayısının yüksek çıkması, farklı özellikleri ölçen maddelerin birbirinden ayırt edilebildiğini gösterirken; puanlayıcı yüzeyi için hesaplanan bu iki değer yüksek olması ise puanlayıcıların bireyleri oldukça farklı puanladıklarını ifade eder (İlhan, 2015). Çok yüzeyli Rasch analizi sonucunda, değişkenlik kaynakları (birey, madde, puanlayıcı, görev, ölçüt, zaman, uygulama vb.) tek bir ölçek üzerine yerleştirilmektedir. Her bir değişkenlik kaynağından araştırmada önemli olduğu düşünülen ikili kombinasyonlar arasındaki etkileşim de incelenebilmektedir. Puanlayıcı davranışını araştırmak aynı zamanda yanlılık hakkında da yorum yapma imkânı verebilmektedir.

Göz İzleme

Göz izleme tekniği, göz bebeği hareketlerinin kızılötesi ışınlar yardımıyla izlenmesi ve kullanıcıların bakış örüntülerinin çıkarılmasıdır (Wotschack, 2009). Göz izleme cihazlarıyla göz bebeği hareketleri izlenerek, çeşitli veriler (metrikler) elde edilir (Lai ve ark., 2013). Elde edilen veriler, araştırmaya katılan bireylerin ekranda nereye, ne kadar süre ve kaç defa baktıkları, dikkatlerini nereye yoğunlaştırdıkları hakkında bilgi verir. Göz hareketleri ölçümleri iki temel üzerine kurulmuştur. Bunlar sabitlemeler ve sıçramalardır. Göz hareketleri metriklerinin hesaplanması için ilk olarak ilgi alanlarının tanımlanmasını gerekir. Tanımlanan farklı ilgi alanlarına ait metriklerin sunumunda tanımlayıcı istatistikler kullanılmaktadır. Bunlar; toplam katılımcı, ortalama, standart sapma, minimum, maximum, toplam ve ortanca değerdir. Göz Hareketi verileri metrik olarak ifade edilmektedir. Göz Hareketi metrikleri; zamansal, uzamsal ve sayım odaklı metrikler olmak üzere üç temel bölüme ayrılmaktadır (Lai ve ark., 2013). Zamansal ölçekte yer alan metrikler, belirli göz hareketlerinde harcanan zamanı gösteren ölçümleri içermektedir. Bunlar; toplam sabitleme süresi, ortalama sabitleme süresi ve ilk sabitlemeye kadar geçen süre olarak ifade edilebilmektedir. Uzamsal ölçekte yer alan metrikler; konumlar, mesafeler, yönler, sabitlemelerin uzunluğu ve sıçramaların genişliği gibi ölçümlerdir. Sayım ölçeğinde ise göz hareketlerinin sıklığı ile ilgili metrikler yer almaktadır. Bunlar toplam sabitleme sayısı, sıçramaların sayısı gibi ölçümlerdir (Lai ve ark., 2013). Toplam sabitleme süresi, izleyicinin ekran üzerinde belirlenen ilgi alanı veya ilgi alanları üzerinde toplam bakış süresi olarak ifade edilebilmektedir. Toplam kayıt süresi, izleyicinin ekran üzerinde toplam bakış süresi olarak ifade edilmektedir. Ekran üzerinde izlenen görevin zorluğu, izleyicinin gösterdiği çabanın boyutu, toplam kayıt süresi incelenerek yorumlanabilmektedir (Jian, 2016; Jian, 2017; Tepgeç ve Seferoğlu, 2019). Ortalama sabitleme süresi, toplam sabitleme süresinin toplam sabitleme sayısına oranı olarak ifade edilmektedir. İzleyicinin gösterdiği çabanın boyutu olarak yorumlanmaktadır (Jian, 2016; Jian, 2017; Tepgeç ve Seferoğlu, 2019). İlk sabitlemeye kadar geçen süre, izleyicinin ekran üzerinde ilk göz hareketi kayıt süresinin başlamasından ilgi alanı üzerinde ilk sabitlemeye

kadar geçen süre olarak ifade edilmektedir. Bu ölçüm değeri, ilgi alanın ne kadar dikkat çekici olduğu hakkında yorum yapma imkânı vermektedir. Ayrıca bilişsel olarak ilk yoğunlaşmanın gerçekleştiği an olarak değerlendirilebilmektedir (Lai ve ark., 2013). İlk sabitleme süresi, ilgi alanı üzerinde ilk odaklanmanın süresi olarak ifade edilebilmektedir. Bu değer ısı haritası üzerinde gözlemlenebilmektedir. Toplam sabitleme sayısı ise ilgi alanı üzerinde her odaklanmanın toplam sayısı olarak ifade edilmektedir. Sabitleme sayısı da yine ilgi alanının dikkat çekiciliğini ve izleyicinin gösterdiği çabanın bir göstergesi olarak ifade edilmektedir. Sabitlemelerin sayısı parametresi okurların dikkat becerisi ile ilgili olarak bilişsel süreçlerine kanıt oluşturabilmektedir (Özer ve ark., 2020).

Göz izleme sonucunda farklı şekillerde görsel veriler elde edilmektedir (Lai ve ark., 2013). Bunlar aşağıdaki gibidir:

1.Sıcaklık Haritaları: Göz hareketlerinin odaklandığı noktaların yoğunluğuna göre farklı renklerle ve renk tonları ile oluşturulmaktadır. En yoğun bölge kırmızı, sonra sarı ve en az yeşil renk olarak görselleştirilmiştir. Sıcaklık haritaları katılımcıların nereye daha çok odaklandığını görsel olarak ortaya koymaktadır.

2.Yol Haritaları: Göz hareketlerinin alanlara odaklanma süresi ve takip ettiği yol görselleştirilmiştir. Katılımcıların nereye ne kadar süre ile baktığı, farklı alanların üzerindeki balonların farklı büyüklükte olmasıyla anlaşılmaktadır. Ekran üzerinde bakışların izlenmiş olduğu yol ise çizgiler yardımıyla görselleştirilmektedir.

3.Kümülatif Haritalar: Bu haritalarda, toplam kayıt süresi içinde göz izleme ile takip edilen alan genel olarak görselleştirilmektedir.

İlgili Araştırmalar

İlgili araştırmalar puanlayıcı ve göz izleme araştırmaları olarak iki ayrı başlık altında verilmiştir.

Puanlayıcı Güvenirliđi ile İlgili Arařtırmalar

Yelbođa ve Tavřancıl (2010) yapmıř oldukları alıřmada, bir iř performans leđinden elde edilen veriler iin KTK ve Genellenebilirlik Kuramına gre gvenirlik katsayılarını karřılařtırmıřtır. Sonu olarak aynı lmler iin KTK ve Genellenebilirlik Kuramı iin ok deđiřkenli modelle elde edilen gvenirlik katsayıları uyumlu olduđu belirtilmiřtir.

Gzkara Bađ ve ark. (2010) tarafında yapılan alıřmada, 2x2 (2 gzlemci 2 lm) sonularının gvenirliđini, Basit Uyum Katsayısı, Cohen Kappa Uyum Katsayısı ve PABAK istatistiklerini arařtırmıřlardır. Arařtırma sonucunda yanlılık ve prevelans etkisini azaltmak iin PABAK istatistiđinin kullanılmasını nermiřlerdir.

Pekin (2015) , puanlayıcılar arası gvenirliđi farklı kuramlarla karřılařtırmıřtır. Otizm Sosyal Beceriler Profili leđi, beř puanlayıcı tarafından tamamlanmıř ve elde edilen puanlar, KTK ve Genellenebilirlik Kuramına gre analiz edilmiřtir. Puanlayıcıların i tutarlılıđı Cronbach Alfa Katsayısı, puanlayıcılar arası gvenirlik Kendall Uyum Katsayısı ile, puanlayıcılar arası korelasyon katsayısı ve puanlar arasındaki iliřki, iliřkili rneklemelerde varyans analizi ile arařtırılmıřtır. Genellenebilirlik Kuramına gre tasarlanan farklı desenlerle arařtırma tamamlanmıřtır. Sonu olarak, Genellenebilirlik Kuramından elde edilen verilerin, KTK' ya gre daha detaylı bilgi verdiđi belirtilmiřtir. Puanlayıcılar arası uyum dzeyi her iki model de de dřk dzeyde olduđu belirtilmiřtir.

İlhan (2015) tarafından yapılan alıřmada, puanlayıcıların aık ulu matematik soruları iin standart ve SOLO taksonomisine dayalı rubriklere verdikleri puanlara ok yzeyli Rasch Modeli uygulanmıřtır. 7 matematik đretmeni puanlayıcı, 104 ortaokul đrencisinin matematik sorularına verdikleri cevapları, standart ve SOLO taksonomisine dayalı rubriklerle puanlamıřlardır. Puanlar ok yzeyli Rasch modeline gre incelenmiřtir. Bunun yanında puanlayıcılara standart rubriklerle ilgili dřnce anketi ve SOLO taksonomisine dayalı rubriklerle ilgili dřnce anketi uygulanmıřtır. Puanlayıcıların bu anket maddelerine verdikleri cevaplar, kapalı ulu maddeler iin aritmetik ortalama, aık ulu maddeler iin betimsel analiz

yöntemiyle analiz edilmiştir. Her iki rubrik için, yetenek düzeyi farklı öğrenciler için ayırt ediciliği yüksek ve madde güçlük düzeyleri arasında anlamlı bir farklılık olduğu belirtilmiştir. Standart rubrik kullanılarak yapılan puanlamalarda, merkeze yönelme etkisi ve puanlayıcı x madde yanlılığı olduğu, katılık-cömertlik açısından anlamlı bir fark olduğu, puanlayıcı güvenilirliğinin düşük olduğu belirtilmiştir. SOLO taksonomisine dayalı rubrik kullanılarak yapılan puanlamalarda, halo etkisi, tutarsızlık, merkeze yönelme, katılık-cömertlik, yanlılık gibi özelliklerin gözlemlenmediği belirtilmiştir. Bunun yanında, puanlayıcı güvenilirliğinin yüksek olduğu ve puanlayıcıların benzer katılık- cömertlikte puan verdikleri belirtilmiştir.

Wang ve ark. (2017) araştırmalarında, öğrencilerin okuduğunu anlama ve yazma becerilerinin birlikte değerlendirilmesine imkân veren, sentez yazım makalelerine (integrated writing essays) puanlayıcıların vermiş oldukları puanların uyumu, puanlayıcı algıları Rasch analizi ve karma ölçme modeline göre incelemişlerdir. Nitel analiz sonucunda puanlamaların uyumsuzluğu değerlendirilen yazının uzunluğu, yazının orijinalliği ve yazının odak noktasının kavranmasında puanlayıcıların farklı düşüncelere sahip olmasından kaynaklandığını vurgulamışlardır.

Bıkmaz Bilgen ve Doğan (2017) yapmış oldukları çalışmada, puanlayıcılar arası uyumu farklı istatistiklerle araştırmış ve sonuçları karşılaştırmışlardır. Puanlayıcı sayısı ve puanlama anahtarı türü değiştirilerek farklı güvenilirlik analizleri karşılaştırılmıştır. Puanlayıcı sayısı arttıkça güvenilirlik düştüğü, amacına göre farklı güvenilirlik istatistiklerinin kullanılması gerektiği sonucuna varıldığı ifade edilmiştir.

Seheryeli Yıldırım (2018) tarafından yapılan çalışmada, yazılı anlatım becerisini, analitik puanlama anahtarı ile elde edilen ölçümlerin, farklı istatistiksel kuramlarla analiz edilerek güvenilirliği nasıl etkilediği araştırılmıştır. Öğrencilerin hazırladığı kısa hikâyeler, 11 maddeden oluşan puanlama anahtarı ile 7 puanlayıcı tarafından puanlanmıştır. Elde edilen veriler KTK, MTK ve Genellenebilirlik Kuramına göre incelenmiştir. Sonuç olarak, hata varyansı kaynaklarını açıklamada ve güvenilirlik belirlemede Genellenebilirlik Kuramının, madde bazında hata kestirimlerinde ve yetenek düzeyi belirlemede Madde Tepki Kuramının,

Klasik Test Kuramı'ndan daha ayrıntılı bilgiler ortaya koyduğu belirtilmiştir. Puanlayıcılar arası güvenilirlik analizi sonuçlarına göre KTK ve Genellenebilirlik Kuramı arasında anlamlı bir fark olduğu, KTK ile MTK arasında anlamlı bir fark olmadığı belirtilmiştir.

Engelhard ve ark., (2018) yapmış oldukları çalışmada, psikometrik ve bilişsel yaklaşımla puanlayıcılar arasındaki uyumu incelemiştir. Psikometrik bakış açısını Rasch Modeli temsil ettiği belirtilmiştir (Rasch,1960; 1980). Bilişsel bakış açısını ise puanlayıcı doğruluk modelinin temsil ettiği ifade edilmiştir (Engelhard, 1996,2013). Yedinci sınıf öğrencilerinin makaleleri, analitik puanlama aracına göre puanlayıcılar tarafından değerlendirilmiştir. Hem bilişsel hem psikometrik olarak puanlayıcıların uyumunun incelenmesi sonucunda değerlendirme kalitesini artırmak için puanlayıcı eğitiminin önemi vurgulanmıştır.

Yılmaz (2019) araştırmasında, 12 maddeden Oluşan Zarar Verici Sosyal Davranışlar Ölçeğini, 20 otizmlili bireye, 2 puanlayıcı ile uygulamıştır ve sonuçlar Klasik Test Kuramı ve Genellenebilirlik Kuramına göre incelemiştir. Sonuç olarak hangi kuramın daha fazla bilgi verdiği yorumlanmıştır. Uyum değerleri Cronbach Alfa, Kendall Uyum Katsayısı ve Person Momentler Çarpımı ile elde edildiği belirtilmiştir. Puanlayıcılar arası ilişki, ilişkili örneklemelerde t-testi ile incelenmiştir. Genellenebilirlik Kuramına göre çaprazlanmış ve yuvalanmış iki desen oluşturulmuş ve buna göre karar çalışması yapıldığı belirtilmiştir. Değişkenlik kaynaklarının bir arada yer alması nedeniyle Genellenebilirlik Kuramının daha bilgi verici olduğu ve kullanılmasının daha yararlı olacağı belirtilmiştir.

Arslan Mancar (2019) çalışmasında, farklı sayıdaki puanlayıcının verdikleri puanların tutarlılığını KTK ve genellenebilirlik kuramına göre karşılaştırmıştır. Puanlayıcıların analitik puanlama anahtarı kullanarak vermiş oldukları puanlar kullanılmıştır. İki, üç ve beş puanlayıcının puanları karşılaştırılmıştır. Kappa ve Krippendorff Alfa Katsayısı kullanılarak uyum düzeyleri belirlenmiştir. Genellenebilirlik kuramına göre tümüyle çaprazlanmış desene göre, puanlayıcı sayısı değişimiyle karar çalışması sonucunda elde edilen G ve Phi

katsayıları ile güvenilirlik analizleri yapılmıştır. Sonuç olarak puanlayıcı sayısının artırılmasının güvenilirliğe katkı sağladığı belirlenmiştir.

Dođru (2019) alıřmasında, karma testlerin psikometrik zelliklerini belirlemede KTK ve Rasch analizi sonularını karřılařtırmıřtır ve hangisinin daha gl olduđunu belirlemeye alıřmıřtır. Sonu olarak geniř lekli testlerde Rasch modelinin KTK'ya iyi bir alternatif olabileceđi ortaya konulmuřtur.

řata (2019) alıřmasında, puanlayıcı gvenirliđini etkileyen problemlili puanlayıcı davranıřlarının belirlenmesi ve arařtırmacı tarafından tasarlanan bir puanlayıcı eđitimi ile problemlili puanlayıcı davranıřlarının zerinde nasıl bir etkisinin olduđunu arařtırmıřtır. Arařtırma n test son test deney ve kontrol grubu ile yarı deneysel bir desen ile oluřturulmuřtur. Analitik dereceli puanlama anahtarı ile toplanan veriler ok yzeyli Rasch Modeli, Genellenebilirlik Modeli ve t testi ile analiz edilmiřtir. Arařtırma sonucunda puanlayıcı eđitiminin gvenirliđe olumlu bir etkisi olduđu yorumu yapılmıřtır.

Tobař (2020) alıřmasında, puanlayıcı yanlılıklarının ve puanlayıcı gvenirliđinin ok yzeyli Rasch lme modeli (YRM) ile belirlenmesini arařtırmıřtır. Arařtırmaya puanlayıcı eđitimi alan 11 đretmen ve 57 ortaokul đrencisi katılmıřtır. đrencilerin hazırlamıř olduđu ikna edici metinler puanlayıcılar tarafından analitik dereceli puanlama anahtarı ile puanlanmıřtır. Elde edilen puanlamalar KTK, GK ve MTK kuramına gre puanlayıcı gvenirliđi incelenmiřtir. đrencilerin yazma becerilerinin deđerlendirilmesinde puanlayıcı katılıđı ve cmertliđi davranıřı gzlemlenmiř ve bireysel dzeyde puanlayıcı yanlılıđı yapıldıđı sonucuna ulařılmıřtır. KTK, GK ve MTK kuramına gre elde edilen farklı puanlayıcı gvenirlik katsayıları yksek ve tutarlı bulunmuřtur. ok yzeyli Rasch lme modeli puanlayıcı gvenirliđinin ve farklılařan puanlayıcı davranıřlarının belirlenmesi iin kullanılabilir alternatif bir lme modeli olabileceđi belirtilmiřtir.

Yakut (2021) alıřmasında, fizyoterapi ve rehabilitasyon blm đrencilerinin eklem hareket lm becerilerini, Genellenebilirlik Kuramına gre deđerlendirmiřtir. Altı farklı deđerlendirmecinin genel izlenim ile performans deđerlendirmesi ve dereceleme leđi ile

değerlendirme sonuçları incelenmiştir. Test tekrar test güvenilirliği ve iki farklı puanlamanın puanlayıcılar arası ilişkisi incelenmiştir. Dereceleme ölçeği puanlama sonuçlarından elde edilen veriler; birey, puanlayıcı ve madde çapraz desenin oluşturulduğu Genellenebilirlik Analizi ile araştırılmıştır. Sınıf İçi Korelasyon Katsayısı ile dereceleme ölçeği puanları ve genel izlenim ile performans değerlendirme ölçeği sonuçları araştırılmıştır. Sonuç olarak, test tekrar test güvenilirliği yüksek bulunduğu belirtilmiştir. Puanlayıcılar arası sınıf içi korelasyon katsayısı yüksek bulunduğu ifade edilmiştir. Fakat puanlayıcılar arasında farklılık olduğu belirtilmiştir. Karar çalışmasında maddelerin puanlamayı etkilemediği, birey ve puanlayıcının daha etkili olduğu yorumu yapılmıştır.

Öksüzoğlu (2022) araştırmasının amacının, 8.sınıf öğrencilerinin Bloom taksonomisine göre üst düzey düşünme becerilerini ölçen, matematik dersi kazanımlarına göre belirlenen, açık uçlu ve çoktan seçmeli maddelere verdiği cevapların, dört puanlayıcı tarafından puanlama anahtarlarına göre puanlaması sonucunda puanlayıcılar arası güvenilirliğin belirlenmesi olduğunu ifade etmiştir. Seçilen PISA sorularının Bloom taksonomi düzeyleri ve 5., 6., 7. ve 8. sınıf matematik dersi kazanımlarına uygun, aynı soru köküne sahip açık uçlu ve çoktan seçmeli sorular belirli aralıklarla ortaokul 8.sınıf öğrencilerine uygulanmıştır. 4 puanlayıcı, hazırlanan analitik puanlama anahtarı ve bütünsel puanlama anahtarına göre öğrencilerin cevaplarını puanladıkları belirtilmiştir. Elde edilen puanların güvenilirlik kestirimleri GK ve KTK ya göre yapıldığı belirtilmiştir. Öğrencilerin açık uçlu ve çoktan seçmeli sorulara verdikleri yanıtlar arasında düşük bir korelasyon elde edildiği vurgulanmıştır. Puanlayıcıların vermiş oldukları puanlar arasındaki tutarlığın yüksek düzeyde olduğu sonucuna varılmıştır.

Gölcür (2022) yapmış olduğu çalışmada, çok yüzeyli Rasch modeli ile puanlama desenlerine göre açık uçlu maddelerin puanlayıcılar arası güvenilirliklerinin karşılaştırılması amaçlanmıştır. Matematik başarı testinde açık uçlu soruları değerlendirmek için hazırlanan, 11 maddeden oluşan, Bütünsel Dereceli Puanlama Anahtarı ile 137, 8.sınıf öğrencisinden elde edilen sonuçlara göre puanlayıcılar arası güvenilirlik, çok yüzeyli Rasch modeline göre

farklı puanlama desenleri karşılaştırılmıştır. Farklı puanlama desenleri güvenilirlik ve kullanılabilirlik açısından karşılaştırılmıştır. Sonuç olarak, çok yüzeyli Rasch modeli puanlama desenlerinden, tamamen çaprazlanmış puanlama deseni (Desen A), tamamlanmamış-bağlantılı puanlama deseni (Desen B ve C), tamamlanmamış-bağlantılı olmayan puanlama deseni (Desen D), tamamlanmamış-bağlantılı sarmal puanlama deseni (Desen E), sarmal puanlama desenine (Desen F) göre analiz yapıldığı belirtilmiştir. Puanlayıcıların katı puanlama yaptığı, katılık-cömertlik davranışları arasında istatistiksel olarak anlamlı fark olduğu, puanlamaya puanlayıcıdan kaynaklanan hataların karıştığı ve Madde 6'nın güçlüğünün en yüksek olduğu belirtilmiştir. Tüm desenlerde madde güçlüğü yüksek ve düşük madde gruplarının, logit değerlerinin farklılaşabildiği belirtilmiştir. Puanlayıcıların, puanlama anahtarı ölçütleriyle tutarlı puanlama yaptıkları ve puanlayıcılar arası güvenilirliğin (uyumun) yüksek olduğu belirtilmiştir.

Göz İzleme Tekniği ile İlgili Çalışmalar

Winke ve Lim (2015) tarafından yapılan çalışmada, 9 eğitimli puanlayıcının, çok kullanılan, 5 alt boyutlu analitik dereceli puanlama anahtarı ile 40 İngilizce makalesini değerlendirirken göz hareketleri incelendiği belirtilmiştir. Göz hareketleri ile bilişsel yetenek arasında bir bağlantı olduğu varsayımı ifade edilmiştir. Bu varsayımdan yola çıkarak, puanlayıcıların analitik dereceli puanlama anahtarı üzerindeki göz hareketleri ile bilişsel yetenekleri ve düşünceleri hakkında fikir edinilebileceği belirtilmiştir. Alt boyutların güvenilirliği sınıf içi korelasyon katsayısı ile belirtilmiştir. Puanlayıcıların dikkati, her bir alt boyutta kontrol edilen kelime sayısı, toplam göz sabitleme süresi ve toplam göz hareketi sayısı ile belirlenmeye çalışılmıştır. Göz izleme ile puanlayıcıların dikkatini belirlediği düşünülen değerler, puanlayıcılar arası güvenilirlikle ilişkilendirilerek, yorum yapılmıştır. Birbiriyle tutarlı puan veren puanlayıcıların, alt boyutlarda ortak dikkat odaklarına sahip olduğu belirtilmiştir. Puanlayıcılar arasındaki anlaşmazlıklar, göz hareketi metrikleri ısı haritalarında doğrudan gözlemlendiği ifade edilmiştir. Analitik puanlama anahtarı üzerinde, puanlayıcıların ilk

okuduđu iki alt boyuta daha fazla dikkat ettikleri, bunun yanında son alt boyuta en az dikkat ettikleri belirtilmiřtir.

Gerçek ve ark. (2016) tarafından yapılan çalışmada, sigara paketi üzerinde bulunan sigaranın sađlıđa zararları ile ilgili görsel uyarılara verilen tepkinin Göz İzleme Yöntemiyle incelendiđi belirtilmiřtir. Sigara içen ve sigara içmeyen katılımcıların, ekran üzerinde sigara paketleri üzerinde bulunan sigaranın sađlıđa zararları ile ilgili görsel uyarılara verdiđi tepkiler göz izleme cihazıyla kaydedildiđi belirtilmiřtir. Daha sonra katılımcılara geri çağırma ve etki formunu tamamlamaları istenmiřtir. Elde edilen sonuçlar cinsiyete göre, sigara içen ve sigara içmeyen bireylere göre görsel uyarının çeřitlendirilmesi gerektiđi belirtilmiřtir. Göz izleme yöntemiyle elde edilen veriler görsel uyarıların yeniden tasarlanması için bir kaynak olarak kullanılabileređi belirtilmiřtir.

Dessus ve ark. (2016), öğrencilerin hangi ipuçlarını algıladıđı ve bu ipucunu nasıl yorumladıđını arařtırmıřlardır. Sınıf yönetiminde, öğrencilerin ihtiyaçlarının algılanması ve bu ihtiyaçların nasıl karşılanacađına iliřkin stratejilerin geliştirilmesinin önemi belirtilmiřtir. Matematik dersinde, 2 uzman öğretmen ve 2 stajer öğretmenin göz izleme metrikleri kaydedilmiřtir. Uzman ve stajer öğretmenlerin göz izleme metrikleri karşılaştırılmıř ve uzman öğretmenler ile stajer öğretmenlerin aralarında farklılık olduđu belirtilmiřtir.

Ballard (2017) arařtırmasında, deđerlendiricilerin rubrik kullanırken davranıřlarını Göz İzleme Tekniđi ile incelemiřtir. Aynı belgeyi puanlayıcıların farklı deđerlendirmesi, hatta aynı puanlayıcının ikinci deđerlendirmede farklı puanlar vermesi puanlayıcı ve puanlama anahtarına olan güvenilirliđi düşürdüđu yorumu yapılmıřtır. Bu arařtırma kapsamında puanlama yaparken puanlama anahtarında nelere dikkat edildiđi, belgeye bakarken o andaki dikkatin nerelere yoğunlařtıđı yorumlanmaya çalışılmıřtır. Puanlayıcı eğitimi verildikten sonra puanlayıcılar, dereceli puanlama anahtarına göre makale puanlama görevlerini tamamlamıřlardır. Beř hafta sonra alternatif bir dereceli puanlama anahtarına göre tekrar aynı makaleler puanlayıcılar tarafından puanlanmıřlardır. Puanlayıcıların dereceli puanlama anahtarına bakarken göstermiř oldukları davranıřlar yorumlanmıřtır.

Gümüşbaş (2019) tarafından yapılan çalışmada, hakemlerin görsel dikkatlerinin göz takip sistemi ile incelenmesi ve sabitleme sayıları, süreleri ve yerleri belirlenerek sporcuların ve hakemlerin performanslarına katkı sağladığı belirtilmesidir. Taekwondo sporunda yer alan 18 hareketin 1 doğru 2 yanlış fotoğrafı ekran üzerinden hakemlere izletilerek, göz takip cihazı ile göz hareketleri kayıt altına alınmıştır. Hakemlerin göz hareketi ölçümleri, doğru ve yanlış fotoğraflar için ortalama sabitleme sayıları, ortalama sabitleme süreleri incelenmiştir. Doğru ve yanlış hareketin yer aldığı fotoğraflar arasında ortalama sabitleme süreleri ve sayıları açısından anlamlı bir fark olduğu belirtilmiştir. Baş bölgesindeki hataların daha az tespit edildiği ve bu bölgede sabitlemelerin daha az olduğu bulgusuna ulaşıldığı ifade edilmiştir.

Coşkun (2019) , giyilebilir göz takip cihazı ile sınıf yönetiminde görsel dikkat gelişimi incelemiştir. Göz takip cihazını kullanan öğretim üyelerinin aynı zamanda video kayıtları alındığı belirtilmiştir. Üç hafta süreyle devam eden çalışmada, her ders sonunda göz izleme verileri ve video kayıtları sonucuna göre, araştırmaya katılan üniversite öğretim üyeleri ile sesli düşünme oturumları gerçekleştirilmiştir. Bu çalışmanın sonucunda, göz izleme yönteminde elde edilen metriklerin, görsel algı hakkında bilgi verdiği, görsel dikkatin geliştirilmesinde rol oynadığı belirtilmiştir.

Nakhaeizadeh ve ark.(2020) tarafından yapılan çalışmada, iskelet kalıntılarının değerlendirilmesinde göz izleme teknolojisi bir araç olarak kullanılabileceği belirtilmiştir. Giyilebilir göz izleme teknolojisi ile 3 uzman değerlendirmecinin, iskeletlerin ölüm yaşı ve cinsiyetlerini belirlerken, göz izleme metrikleri incelenmiştir. Uzman değerlendirmecilerin sabitleme noktaları, sabitleme sayıları ve toplam bakış süreleri incelenmiştir. Elde edilen veriler değerlendirmecilerin, değerlendirme yaparken kaydedilen göz hareketlerinin izledikleri yol hakkında bilgi verdiği sonucuna ulaşılmıştır.

Wang ve ark. (2020) tarafından yapılan çalışmada, kısa eğitici videolarda ipuçlarının öğrenme süreçleri ve çıktıları üzerindeki etkisini incelemişlerdir. Dört farklı koşulda (ipucu yok, metinsel ipucu, görsel ipucu, birleşik hem metinsel hem görsel ipucu) hazırlanan

videolar, öğrencilere rastgele olarak verilmiştir. Öğrenciler, videoları izlerken göz takip cihazıyla, göz hareketleri incelenmiştir. Öğrencilerin öğrenme çıktıları, kalıcılık ve aktarım testleri ile ölçülmüştür. Sonuçlar incelendiğinde ise öğrencilerin, odaklanmalarına ve başarılarının artmasına ipuçlarının katkıda bulunduğu belirtilmiştir. Ayrıca görsel ve birleşik (metinsel +görsel) ipuçları, öğrencilerin transfer testlerinde, bilgiyi organize etmede ve bütünleştirmede daha başarılı olmalarına yardımcı olduğu sonucuna varılmıştır.

Özmen (2021) araştırmasının amacı, gerçek bir çevrimiçi alışveriş sürecinde kullanıcıların web arayüz etkileşimlerinde duygu durumlarındaki değişimleri incelemektir. Bu amaç kapsamında katılımcıların bir e-ticaret sitesindeki alışveriş süreci boyunca göz hareketleri, EEG, GSR, Nabız ve yüz ifadelerinde demografik özelliklere ve web sayfası bileşenlerine göre değişiklik olup olmadığı sorusuna cevap aranmıştır. Elde edilen bulgulara göre katılımcılar, çevrim içi alışveriş sürecinde web sayfası arayüzü ile ilgili ürün arama, renk, tasarım, kategorilendirme, filtreleme ve sıralama gibi unsurlara odaklandıkları belirtilmiştir. Ayrıca katılımcıların Nabız, GSR ve EEG verileri açısından elde edilen duygu durumu değişimlerinde önemli farklılıklar bulunduğu ifade edilmiştir. Arayüz bileşenleri ve demografik faktörler açısından da anlamlı ilişkiler olduğu vurgulanmıştır.

Friedman ve ark. (2022) çalışmalarında, göz izleme verilerinin sınıflandırılmasında, puanlayıcıların etkilerini, üç farklı veri setiyle incelemiştir. Manuel göz hareketi sınıflamaları, otomatik algoritmalar geliştirme aşamasında, temel olarak kullanıldığı ifade edilmiştir. Fakat güvenilirliği düşük ve tutarsız puanlamaların gelişmeye katkı sağlamayacağı belirtilmiştir. Bu nedenle göz hareketi verilerinin doğruluğu ve güvenilirliğini hangi faktörlerin etkilediğinin araştırılması gerektiği belirtilmiştir. Üç farklı veri setinde Puanlayıcılar arası uyum, Cohen Kappa ve F1 Etkinlik Düzeyi analiziyle araştırılmıştır. Sonuçlar incelendiğinde, puanlayıcılar arası güvenirliliğin; göz izleme kalitesi, baş hareketi, kaydedilen konuların özellikleri, puanlama aracının özellikleri ve puanlayıcı eğitimi gibi faktörlerden etkilenebildiği belirtilmiştir.

Coşkuner (2022) tarafından yapılan çalışmada, Göz İzleme Tekniği ile elde edilen metriklerin, test ve madde istatistikleri ile ilişkisi incelenmiştir. İlk olarak 60 çoktan seçmeli madde içeren başarı testi 4848 öğrenciye uygulanmıştır. Testin madde istatistikleri hesaplanmıştır. 60 öğrenci çoktan seçmeli testi doldururken, giyilebilir göz takip cihazıyla göz hareketleri kaydedilmiştir. Ortalama sabitleme süresi metriği ile madde ayırt edicilik indeksi dışındaki bütün madde istatistikleri anlamlı bir ilişki gösterdiği yorumu yapılmıştır. Göz izleme yöntemiyle elde edilen metriklerin, başarı testinden elde edilen test puanları yordayıcısı olabileceği belirtilmiştir. İlk sabitlemeye kadar geçen süre ile toplam ziyaret süresi metriklerinde kızlar lehine anlamlı bir farklılık olduğu belirtilmiştir. Ortalama ziyaret süresi metriğinde erkekler lehine anlamlı bir farklılık bulunduğu belirtilmiştir. Ayrıca zaman odaklı metrikler ile madde istatistikleri arasında güçlü bir ilişki olduğu belirtilmiştir.

İlgili araştırmalar genel olarak değerlendirildiğinde; puanlayıcılar arasındaki uyumun incelenmesinde birden fazla istatistiksel modelin birlikte kullanıldığı çalışmaların yaygın olduğu gözlemlenmektedir. Puanlayıcılar arasındaki uyumsuzluğun nedenleri araştırılmış ve puanlayıcı eğitiminin verilmesinin bu uyumsuzluğu azaltabileceği ifade edilmiştir. Göz izleme yöntemiyle ilgili çalışmalar farklı alanlarda çeşitlilik göstermekle birlikte ilk çalışmalar kullanışlılıkla ilgili yapılmıştır. Göz izleme yönteminin kullanıldığı çalışmalarda son yıllarda bir artış gösterdiği gözlemlenmektedir. Puanlayıcı uyumunun belirlenmesi ile ilgili göz izleme çalışmalarının en güncel araştırmalar olduğu gözlemlenmiştir. Puanlayıcılar arası uyumun göz izleme yöntemiyle araştırıldığı çalışmalar literatürde yer almaktadır. Bunun yanında puanlayıcılar arası uyumun klasik test kuramı, genellenebilirlik kuramı, çok yüzeysel Rasch modeli ve göz izleme metriklerinin birlikte incelendiği bir çalışmaya rastlanmamıştır.

Bölüm 3

Yöntem

Araştırmanın Türü

Araştırmanın yöntemi farklı tekniklerle elde edilen bilgilerin ilişkilendirilmesi nedeniyle ilişkisel, farklı tekniklerle elde edilen bilgilerin değerlendirilmesi nedeniyle betimsel çalışmalar arasında sayılabilir. Araştırma betimsel ve karşılaştırmalı araştırma içerisinde değerlendirilebilir (Karasar, 2000; Büyüköztürk ve ark., 2016).

Çalışma Grubu

Araştırmanın katılımcıları, sınıf içi performans sergileyen öğrenciler ile öğrencilerin performansını değerlendiren puanlayıcılardan oluşmaktadır.

Öğrenci Grubu

Araştırma 2018-2019 Eğitim-Öğretim Yılı Bahar Dönemi'nde Hacettepe Üniversitesi Eğitim Fakültesinde verilen Eğitim Felsefesi dersini alan öğrenci grubu ve Eğitim Bilimleri Enstitüsü Eğitimde Ölçme ve Değerlendirme Doktora Programında Ölçmenin Psikolojik Temelleri dersini alan öğrenci grubu, üzerinden yürütülmüştür. Dersin öğretim üyesi, çeşitli konular ile ilgili sınıfta sunum performansı göstermeleri amacıyla öğrencileri gruplara ayırmıştır. Bu gruplar arasından "Yaraticılık", "İdealizm" ve "Realizm" konuları üzerinde sunum yapan 3 öğrenci, araştırmanın öğrencilere ilişkin çalışma grubunu oluşturmaktadır.

Puanlayıcı Grubu

Araştırmada, sınıf içi uygulamalarında ve Göz izleme yöntemiyle ekran üzerinde tamamlanan değerlendirmeye, Eğitimde Ölçme ve Değerlendirme alanında doktora derecesine sahip beş uzman katılmıştır. Uzmanlardan ikisi pilot uygulamadan sonra çeşitli nedenlerle çalışmaya devam edememiştir. Bu nedenle nihai araştırma üç uzmanın değerlendirmelerine göre raporlanmıştır.

Veri Toplama Süreci

Uygulama için ilk aşamada öğrenci sunumları videoya alınmıştır. Bunun yanında, üç puanlayıcı sunum esnasında hazır bulunarak sınıf içinde değerlendirmelerini, uzman görüşleri alınarak hazırlanmış sunum değerlendirme dereceli puanlama anahtarına göre tamamlamışlardır. Bir hafta sonra Tobii Pro X2-60 cihazı kullanılarak, puanlayıcılar öğrencilerin sunum videolarını bilgisayar aracılığıyla izlemiş ve sunum değerlendirme dereceli puanlama anahtarına göre tekrar puanlamışlardır. Bu esnada puanlayıcıların, göz hareketleri bilgisayara eklenen Tobii Göz İzleme Cihazı ile kayıt altına alınmıştır. Puanlayıcıların sunum değerlendirme dereceli puanlama anahtarına göre sınıf içinde ve ekran üzerinde verdiği puanlar göz izleme tekniği ile puanlayıcıların kayıt altına alınan göz hareketleri çalışmanın ana verilerini oluşturmaktadır. Sınıf İçi uygulaması için belirlenen gruplar, öğrencilerin gruplar halinde sunum performanslarını gösterdikleri Eğitim Fakültesi Lisans Programı, 2.sınıf dersi Eğitim Felsefesi dersinde idealizm ve realizm sunumları ve Eğitim Bilimleri Enstitüsü Eğitimde Ölçme ve Değerlendirme doktora Programı Ölçmenin Psikolojik Temelleri dersi yaratıcılık ve yaratıcılığın ölçülmesi sunumu olarak belirlenmiştir.

Pilot çalışma

Uygulama öncesinde, 2018-2019 güz döneminde Eğitim Fakültesi Sınıf Öğretmenliği laboratuvar uygulamaları dersi kapsamında sınıf içi grup çalışmasında video kullanılarak uygulama yapılmıştır. Sunum değerlendirme dereceli puanlama anahtarı kapsamında, sınıf içinde uygulama yapan gruplar, dersin görevli öğretim elemanı ve puanlayıcılar tarafından puanlanılmışlardır. Öz değerlendirme, akran değerlendirme ve 5 uzman puanlayıcı değerlendirmesinin yer aldığı 4 farklı grubun değerlendirme sonuçları incelenmiştir. Grup çalışmasında ilgi alanı belirleme ve puanlayıcı göz hareketlerinin anlamlı bir sonuç elde edilmesinin zorluğu nedeniyle asıl uygulamada grup çalışması yerine sınıf içi sunum performansının değerlendirilmesinin daha uygun olacağına karar verilmiştir.

Veri Toplama Araçları

Öğrenci gruplarının sunum performansının değerlendirilmesinde kullanılmak üzere bir puanlama aracı geliştirilmiştir. Dereceli puanlama anahtarının geliştirilme aşamasında Lawshe tekniği kullanılmıştır (Yurdugül, 2005). Lawshe tekniğine göre;

- Alan uzmanları grubunun oluşturulması
- Aday ölçek formunun hazırlanması
- Uzman görüşlerinin değerlendirilmesi
- Maddelere ilişkin kapsam geçerlik oranlarının belirlenmesi
- Ölçeğe ilişkin kapsam geçerlik indeksinin bulunması
- KGO ve KGİ değerlerine göre nihai formun oluşturulması

Tablo 3

Uzman Grubunun Özellikleri

<i>Cinsiyet</i>	N
Kadın	22
Erkek	13
<i>Eğitim Düzeyi</i>	
Doktora mezunu	15
Doktora devam ediyor	20
<i>Çalışma Alanı</i>	
Eğitimde Ölçme ve Değerlendirme	23
Eğitim Programları ve Öğretimi	7
Eğitim Yönetimi	5

Dereceli puanlama ölçeğinin geliştirilme aşamasında ilk olarak literatür incelenerek, araştırmanın amacına uygun olarak araştırmacı tarafından sunum becerilerini temsilen 23 ifadeden oluşan bir madde havuzu hazırlanmıştır (Carroll, 2006; Langan ve ark., 2005; Langan ve ark., 2008). Tablo 3 'te özellikleri maddeler halinde belirtilmiş olan uzman grubu, bu madde havuzu 20 eğitim bilimleri lisansüstü öğrencisi ile 16 eğitim bilimleri lisansüstü

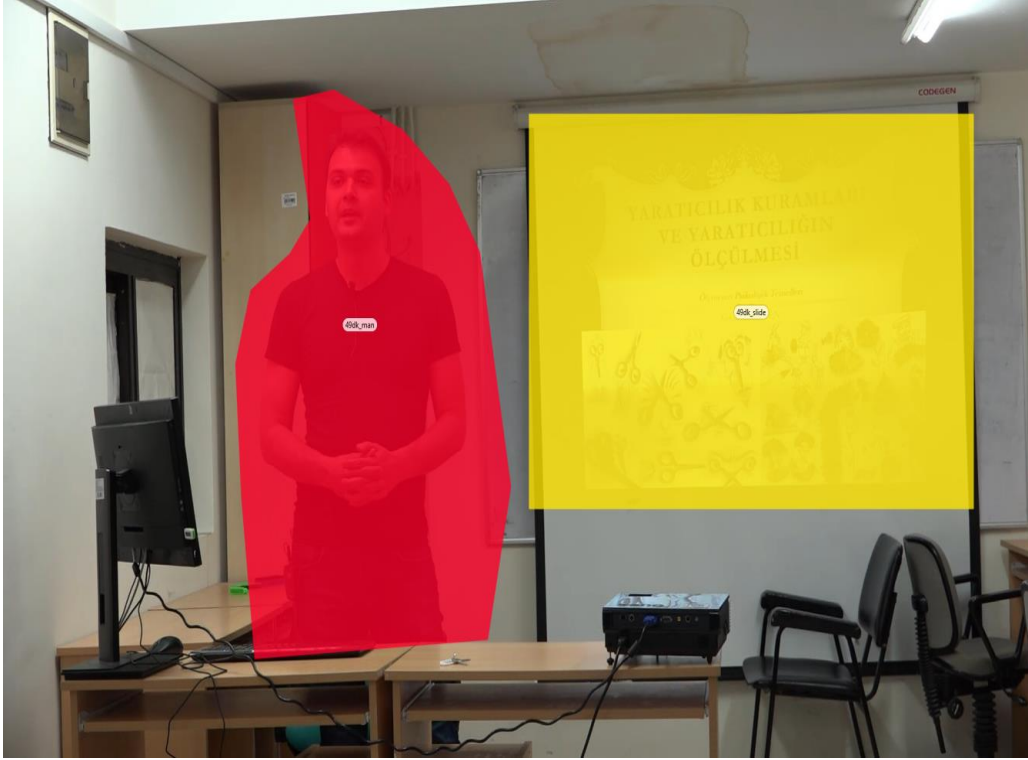
mezunu akademisyene, kapsam ve uygunluğunu değerlendirmeleri üzere gönderilmiştir. Uzmanlar maddeleri uygun, düzeltilmeli ve uygun değil olmak üzere 3 dereceli bir formla puanladıktan sonra elde edilen puanlar incelenmiştir. Her bir maddeye verilen toplam puanlar elde edilmiştir. Kapsam geçerlik oranı (KGO) değeri hesaplanmıştır. KGO değeri her bir maddeye olumlu puan verenlerin, toplam puan verenlerin yarısına oranının bir eksiği olarak hesaplanmıştır. Maddelerin istatistiksel olarak anlamlı olup olmadığını test etmek için kapsam geçerlik ölçütü geliştirilmiştir. Kapsam Geçerlik Ölçütü için standart normal dağılım ilkelerinden yararlanılarak $p=0,05$ anlamlılık düzeyinde KGO değerinin minimum değerleri tabloya dönüştürülmektedir (Veneziano & Hooper, 1997; Yurdugül, 2005). Uzman sayısı $N=35$ değeri için minimum KGO değeri 0,31 olarak belirtilmiştir (Veneziano & Hooper, 1997; Yurdugül, 2005). Her bir madde için KGO değeri 0,31 in altındaki maddeler çıkarılmış ve 0,31 in üstünde bulunan maddelerin ortalama kapsam geçerlik indeksi ($KGİ = 0,76$) olarak belirlenmiştir. $KGİ > KGÖ$ olduğu için ölçeğin kapsam geçerliğinin istatistiksel olarak anlamlı olduğu yorumu yapılabilmektedir. Birden fazla boyutu olan ölçeklerde her bir alt boyut için $KGİ$ değeri hesaplanmaktadır. Değerlendirmeler sonucunda 19 madde uygun bulunmuştur. “Tamamen yeterli”, “yeterli”, “kısmen yeterli” ve “yeterli değil” derecelendirmesi ile 19 maddelik dereceli puanlama anahtarının nihai haline karar verilmiştir. Dereceli puanlama anahtarının iç tutarlılığı yapılan pilot uygulamada $n= 60$ katılımcıyla elde edilen sonuçta Cronbach alfa katsayısı ile incelenmiş ve $\alpha= 0,87$ olarak hesaplanmıştır. EK-A ‘da yer alan sunum değerlendirme dereceli puanlama anahtarı oluşturulmuş ve maddeleri giriş, süreç ve sonuç başlıkları altında sunulmuştur.

Göz İzleme Yöntemi

Göz izleme yöntemi sonucunda puanlayıcıların göz hareketleri, göz izleme cihazı ile kayıt altına alınmıştır. Daha sonra, Tobii Studio programıyla, kayıt altına alınan göz hareketlerinin nasıl düzenleneceği belirlenmektedir. İlk olarak araştırmanın amacına uygun olarak, her bir sunum videosu için ilgi alanları belirlenmiştir.

Şekil 1*İdealizm Sunumu İlgili Alanları*

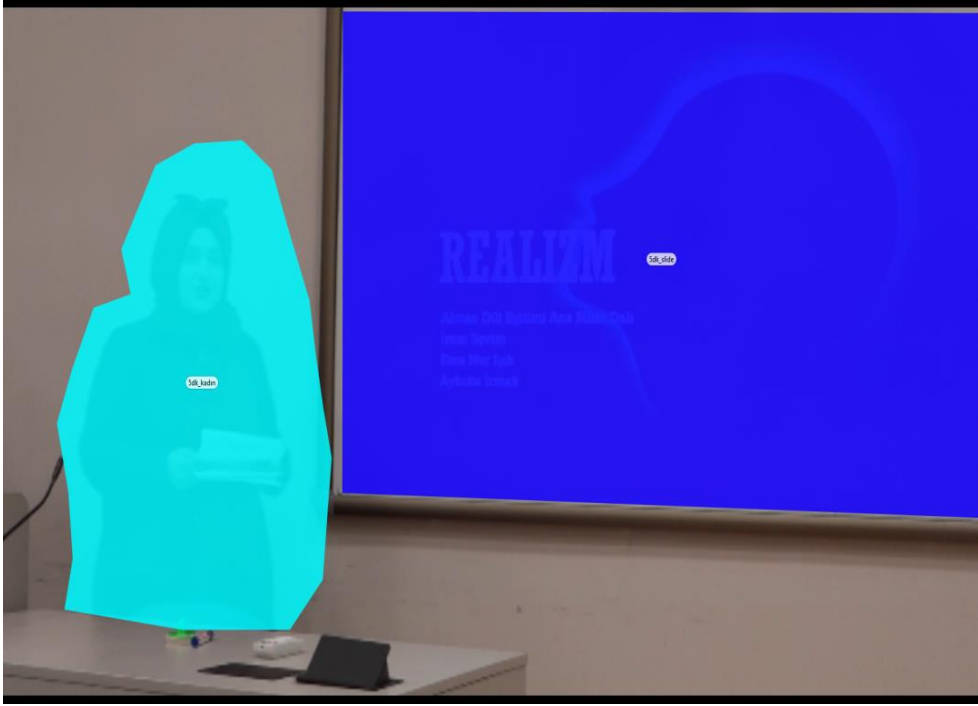
Şekil 1 'de görüldüğü gibi İdealizm sunumu için göz izleme verileri iki ilgi alanı metrikleri dikkate alınarak karşılaştırılmıştır. İLÖ1 mor renk ile taranan alanın içinde kalan göz izleme verilerinin elde edildiği alandır. İLt1 ise kırmızı renk ile taranan alanın içinde kalan göz izleme verilerinin elde edildiği alandır. İdealizm sunumu için puanlayıcıların göz izleme metriklerine ait betimsel istatistikler bu iki alanın karşılaştırmalarını içermektedir.

Şekil 2*Yaratıcılık Sunumu İlgili Alanları*

Şekil 2 'de görüldüğü gibi Yaratıcılık sunumu için göz izleme verileri iki ilgi alanı metrikleri dikkate alınarak elde edilmiştir. İLÖ2 kırmızı renk ile taranan alanın içinde kalan göz izleme verilerinin elde edildiği alandır. İLt2 ise sarı renk ile taranan alanın içinde kalan göz izleme verilerinin elde edildiği alandır. Yaratıcılık Sunumu için göz izleme metriklerine ait betimsel istatistikler bu iki alanın karşılaştırmalarını içermektedir.

Şekil 3

Realizm Sunumu İlgi Alanları



Şekil 3 'de görüldüğü gibi Realizm sunumu için göz izleme verileri iki ilgi alanı metrikleri dikkate alınarak elde edilmiştir. İLÖ3 turkuaz renk ile taranan alanın içinde kalan göz izleme verilerinin elde edildiği alandır. İLt3 ise mavi renk ile taranan alanın içinde kalan göz izleme verilerinin elde edildiği alandır. Realizm sunumu için Puanlayıcıların göz izleme metriklerine ait betimsel istatistikler bu iki alanın karşılaştırmalarını içermektedir.

İlgi alanı belirlendikten sonra elde edilen göz izleme metrikleri aşağıda sıralanmaktadır.

- İlgi alanı üzerinde ortalama sabitleme süresi
- İlgi alanı üzerine gelinceye kadar geçen süre: Bu ölçüm bir test katılımcısının aktif bir ilgi alanı veya ilgi alanı grubuna ilk kez odaklanmasının ne kadar sürdüğünü ölçer.

- İlgi alanı üzerinde gerçekleşen ilk sabitlenmenin süresi: Bir ilgi alanı veya ilgi alanı grubu üzerinde ilk fiksasyon süresidir (saniye). Bu değer ısı haritasıyla gözlemlenebilmektedir.
- İlgi alanı üzerinde toplam sabitlenme süresi: Bir ilgi alanı içindeki veya bir ilgi alanı grubuna ait tüm ilgi alanı içindeki her bir bireysel fiksasyonun süresidir (saniye).
- İlgi alanı sabitlenme sayısı

Bu verilere ek olarak göz izleme sürecinde oluşan yol haritası ve sıcaklık haritası elde edilmiştir.

Verilerin Analizi

Veri toplama araçları puanlayıcıların performans değerlendirme için hazırlanan sunum değerlendirme dereceli puanlama anahtarı ve göz izleme tekniği sonucunda elde edilen puanlayıcı göz izleme verileri olmak üzere iki boyutta ele alınabilir.

Birinci alt problem kapsamında, iki puanlayıcı arasındaki uyumu incelemek için Cohen Kappa, üç puanlayıcı arasındaki uyumu incelemek için Fleiss Kappa uyum istatistikleri hesaplanmıştır.

İkinci alt problem kapsamında, puanlayıcılar arasındaki uyumu incelemek için, sınıf içi korelasyon katsayısı hesaplanmıştır.

Üçüncü alt problem kapsamında, performans değerlendirme sürecinde elde edilen puanlar, genellenebilirlik kuramı ile incelenmiştir. Genellenebilirlik analizi için EduG-6e paket programından yararlanılmıştır.

Dördüncü alt problem kapsamında, performans değerlendirme sürecinde elde edilen puanlar çok yüzeyli Rasch modeli ile incelenmiştir. Sonuçlar, çok yüzeyli Rasch modeline göre Facet (Linacre,2018) programında analiz edilmiştir. Uygulama sonucunda her bir yüzey için uygunluk istatistikleri, parametre tahminleri ve her bir parametreye ait standart hata değerleri üretilmektedir. Parametre tahminlerine ait standart hata değerleri ölçümlerin güvenilirliği, uygunluk istatistikleri ölçümlerin geçerliği hakkında bilgi vermektedir (Barkaoui, 2008).

Beşinci alt problem kapsamında, göz izleme yöntemiyle her bir puanlayıcıdan elde edilen göz izleme metriklerinden ortalama sabitleme süreleri, çok yüzeyli Rasch analizi ile incelenmiştir.

Altıncı alt problem kapsamında, göz izleme sonucunda puanlayıcıların göz hareketleri, Tobii Studio programı ile analiz edilerek göz izleme metrikleri elde edilmiştir. Her bir sunum videosu için üç puanlayıcıya ait göz izleme metriklerinin betimsel istatistikleri, yol haritası ve sıcaklık haritası elde edilmiştir.

Bölüm 4

Bulgular ve Tartışma

Bu bölümde, çalışmanın her bir alt problemlerine ilişkin bulgulara, bulgulara yönelik yorumlara ve ulaşılan sonuçların ilgili literatür desteğiyle tartışmasına yer verilmiştir.

Verilere ilişkin temel analizlere geçilmeden önce ön analizler yapılmıştır. Çalışmada yer alan üç puanlayıcının sınıf içi ve bilgisayar ekranından göz izleme tekniği ile 19 madde ve 3 farklı performans görevi için verdikleri puanlara ait betimsel istatistikleri Tablo 5'te görülmektedir.

Tablo 4*Sınıf İçi Uygulama ve Göz İzleme Uygulaması Sonucunda Üç Puanlayıcıdan Elde Edilen Puanların Betimsel İstatistikleri*

	Sınıf İçi Uygulama									Göz İzleme Uygulaması								
	İdealizm			Yaratıcılık			Realizm			İdealizm			Yaratıcılık			Realizm		
	P1	P2	P3	P1	P2	P3	P1	P2	P3	P1	P2	P3	P1	P2	P3	P1	P2	P3
Ortalama	2,26	1,95	2,21	3,37	3,53	2,84	2,21	2,42	2,63	1,68	1,95	2,05	3,53	3,63	3,05	2,16	2,42	2,42
Standart Sapma	0,93	0,71	0,71	0,50	0,51	0,60	0,71	0,84	0,60	0,75	0,62	0,62	0,51	0,50	0,23	0,90	0,96	0,77
Çarpıklık	0,33	0,07	0,69	0,59	-0,12	0,05	-0,34	-0,36	-1,44	0,62	0,03	-0,03	-0,12	-0,59	4,36	-0,34	-0,17	-0,94
Basıklık	-0,52	-0,77	1,26	-1,86	-2,24	0,04	-0,82	-0,48	1,38	-0,86	0,02	0,02	-2,24	-1,86	19,00	-1,75	-0,86	-0,54
Minimum	1	3	1	1	3	1	1	2	1	1	1	1	3	3	3	1	1	1
Maksimum	4	3	4	4	4	4	3	4	3	3	3	3	4	4	4	3	4	3
Toplam	43	37	42	64	67	54	42	46	50	32	37	39	67	69	58	41	46	46

P1: 1.puanlayıcı; **P2:** 2.puanlayıcı; **P3:** 3.puanlayıcı

Tablo 4' te yer alan betimsel istatistikler incelendiğinde hem sınıf içi uygulamada hem de göz izleme uygulamasında ortalama değerlerin birbirine yakın olduğu gözlemlenmektedir. Çarpıklık ve basıklık değerlerinin çoğunluğu -1 ve +1 aralığında olması, verilerin normal dağılım sergilediğini göstermektedir (Baykul, 1999; Büyüköztürk ve ark., 2013; Çokluk ve ark., 2014). Ortalamaların birbirine yakın olduğu, standart sapmaların yaratıcılık sunumunda diğerlerine göre farklılaştığı görülmektedir.

Birinci Alt Probleme İlişkin Bulgular

Araştırmanın birinci alt problemi, "Performans değerlendirme sürecinde klasik tekniklerle elde edilen puanlayıcılar arası uyum düzeyleri nasıldır?" şeklinde ifade edilmiştir. Bu alt problem doğrultusunda sınıf içi ve göz izleme uygulamalarında uzmanların değerlendirme sonuçları arasında uyum Cohen Kappa ve Fleiss Kappa katsayıları ile incelenmiş ve sonuçlar Tablo 6'da verilmiştir.

Tablo 5
Cohen Kappa Analizi Sonuçları

Sunu	Puanlayıcı İkili	Sınıf İçi Uygulama	Göz İzleme Uygulaması
İdealizm	P1 –P2	0,174	0,509*
	P1 –P3	0,570*	0,530*
	P2 – P3	0,205	0,477
Yaratıcılık	P1 –P2	0,273	0,067
	P1 –P3	0,042	0,095
	P2 – P3	0,141	0,108
Realizm	P1 –P2	0,362	0,316
	P1 –P3	0,149	0,580*
	P2 – P3	0,429	0,301

Puanlayıcılar arasındaki uyum, ilk olarak, Cohen Kappa analizi ile araştırılmıştır. Tablo 5'e göre iki puanlayıcının olduğu durumda 0,042 ile 0,580 arasında değişen kappa değerleri elde edilmiştir. Kappa değerleri incelendiğinde sınıf içi puanlama için sadece bir kappa değerinin istatistiksel olarak anlamlı olduğu, göz izleme puanlaması için ise üç değer istatistiksel olarak anlamlı olduğu görülmektedir ($p < 0,05$). Hesaplanan Kappa değerleri

puanlayıcılar arasında, genel olarak, zayıf ve orta derecede uyum olduğunu göstermektedir (Landis & Koch,1977). En düşük uyumun yaratıcılık sunumunda P1-P3 arasında, en yüksek uyumun ise idealizm sunumunda P1-P3 arasında olduğu gözlenmiştir. Kappa istatistiği puanlayıcılar arası uyumu etkileyen hata kaynakları hakkında bilgi vermemektedir. 3 videonun süreleri birbirinden farklı olduğu için puanlayıcı güvenliği farklılaşmış olabilir. Cohen Kappa, sıralama ölçeğinde ikili puanlayıcı karşılaştırması yaptığı için hata puanını yükseltmektedir. Bu nedenle üç puanlayıcı için uyum katsayıları hesaplanmış ve Tablo 7’de sunulmuştur.

Tablo 6
Üç Puanlayıcı İçin Fleiss Kappa Analiz Sonuçları

Sunu	Sınıf İçi Uygulama	Göz İzleme Uygulaması
İdealizm	0,321	0,530*
Yaratıcılık	0,252	0,250
Realizm	0,393	0,422

Tablo 6’da yer alan Fleiss Kappa değerleri incelendiğinde, puanlayıcılar arasında orta derecede bir uyum olduğu görülmektedir (Landis & Koch, 1977). Ayrıca sınıf içi uygulamaya göre, göz izleme sonuçları için elde edilen değerlerin görece daha yüksek olduğu söylenebilir. Puanlayıcılar arası en yüksek uyum değeri göz izleme ile ekran üzerinden yapılan puanlama için idealizm sunumunda, en düşük uyum değeri ise göz izleme ile ekran üzerinden yapılan puanlama için yaratıcılık sunumunda gözlenmiştir.

Sınıf içi ve göz izleme uygulamalarında ortalama puanlara ilişkin uyum, sınıf içi korelasyon katsayısı ile incelenmiş ve sonuçlar Tablo 8’de verilmiştir.

Tablo 7
Sınıf İçi Korelasyon Analizi Sonuçları

Sunu	Sınıf İçi Uygulama	Göz İzleme Uygulaması
İdealizm	0,736*	0,739*
Yaratıcılık	0,155	0,170
Realizm	0,636*	0,681*

Tablo 7’de yer alan sınıf içi korelasyon katsayıları incelendiğinde, idealizm ve realizm sunumlarında puanlayıcılar arası uyumun iyi olduğu görülmektedir (Fleiss ve ark., 2003).

Yaratıcılık sunumu için puanlayıcılar arasındaki uyumunun ise zayıf olduğu bulunmuştur (Fleiss ve ark., 2003). Ayrıca göz izleme uygulaması sonucunda elde edilen sınıf içi korelasyon değerlerinin sınıf içi uygulamalara göre daha yüksek olduğu gözlenmiştir. Puanlayıcılar arası en düşük uyum sınıf içi uygulamada yaratıcılık sunumunda, en yüksek uyum ise göz izleme ile ekran üzerinden yapılan idealizm sunumunda elde edilmiştir.

İkinci Alt probleme İlişkin Bulgular

Araştırmanın ikinci alt problemi “Performans değerlendirme sürecinde elde edilen puanların genellenebilirlik kuramı uygulamasıyla elde edilen varyans kaynaklarının değeri nasıldır?” şeklindedir. Bu alt problem doğrultusunda, puanlayıcıların, sınıf içi ve göz izleme yöntemiyle ekran üzerinden verdiği puanlar genellenebilirlik modeline göre analiz edilmiştir.

Öğretmen adaylarının sunum becerilerine ilişkin puanlayıcıların sınıf içi uygulama sürecinde verdikleri puanlar ve sınıf içi uygulamaların video kayıtlarının göz izleme sistemiyle puanlayıcılar tarafından tekrar değerlendirilmesi sonucu elde edilen puanların genellenebilirlik kuramına göre dört yüzeyle çaprazlanmış desenle sonuçları elde edilmiştir (PxSxUxM). Veriler üç alan uzmanı puanlayıcı P, üç performans görevi S, iki farklı uygulama (sınıf içi ve göz izleme uygulaması) U, madde M ana etkileri; UP, US, UM, PS, PM, SM, UPS, UPM, USM, PSM, UPSM ise ortak etkileri oluşturmaktadır. Genellenebilirlik analizi sonuçları Tablo 8’de verilmiştir.

Tablo 8*Tamamen Çaprazlanmış (UxPxSxM) Desene ilişkin G Çalışması Sonuçları*

Varyans Kaynağı	df	Kareler Toplamı	Kareler Ortalaması	Varyans	Varyans %
<i>U</i>	1	0,292	0,292	0,000	0,00
<i>P</i>	2	0,988	0,494	0,000	0,00
<i>S</i>	2	103,936	51,968	0,421	39,60
<i>M</i>	18	55,784	3,099	0,119	11,20
<i>UP</i>	2	0,532	0,266	0,000	0,00
<i>US</i>	2	2,357	1,178	0,014	1,40
<i>UM</i>	18	1,819	0,101	0,000	0,00
<i>PS</i>	4	10,538	2,635	0,062	5,80
<i>PM</i>	36	17,567	0,488	0,065	6,10
<i>SM</i>	36	26,620	0,739	0,091	8,50
<i>UPS</i>	4	1,450	0,363	0,004	0,40
<i>UPM</i>	36	6,690	0,186	0,000	0,00
<i>USM</i>	36	10,199	0,283	0,000	0,00
<i>PSM</i>	72	14,240	0,198	0,000	0,00
<i>UPSM</i>	72	20,661	0,287	0,287	27,00
Toplam	341	273,673			100

P: puanlayıcı; S:performans görevi (sunu), U: uygulama, M: madde

Tablo 8'de yer alan G çalışması sonuçlarına göre uygulama, puanlayıcı, sunu ve madde değişkenlik kaynakları incelendiğinde toplam varyansın %39,6'sının sunudan geldiği görülmektedir. Bir başka ifadeyle, sunu ana etkisi için (tek değişkenli modelle yapılan) G çalışmasından kestirilen varyans bileşeni (0,421) toplam varyansın %39,6'sını açıklamaktadır. Bu sonuca göre sunumdan sunuma puanlamanın farklılaştığı söylenebilir. UPSM ortak etkisi toplam varyansın %27'sini açıklamaktadır. Madde ana etkisi için kestirilen varyans bileşeni (0,119) toplam varyansın %11,2'sini, SM ortak etkisi ise toplam varyansın %8,5'ini açıklamaktadır. Puanlayıcı ana etkisinin G çalışması ile kestirilen varyans oranının düşük olması, puanlayıcıların bireyler için yaptıkları puanlamalar arasında farklılık bulunmadığı, puanlamaların tutarlı olduğu şeklinde yorumlanabilir. G katsayısı 0,92 ve Phi katsayısı 0,91 olarak hesaplanmıştır. Sunu ve madde varyansının büyük olması, tüm yüzeylerin ortak etkisinin veya tesadüfi hataların büyük olabileceğinin bir göstergesidir. Uygulama ana etkisi ölçmenin temel konusu (ölçme nesnesi) olarak değerlendirildiğinde, ana değişkenliğin bu yüzeyden gelmesinin beklendiği söylenebilir.

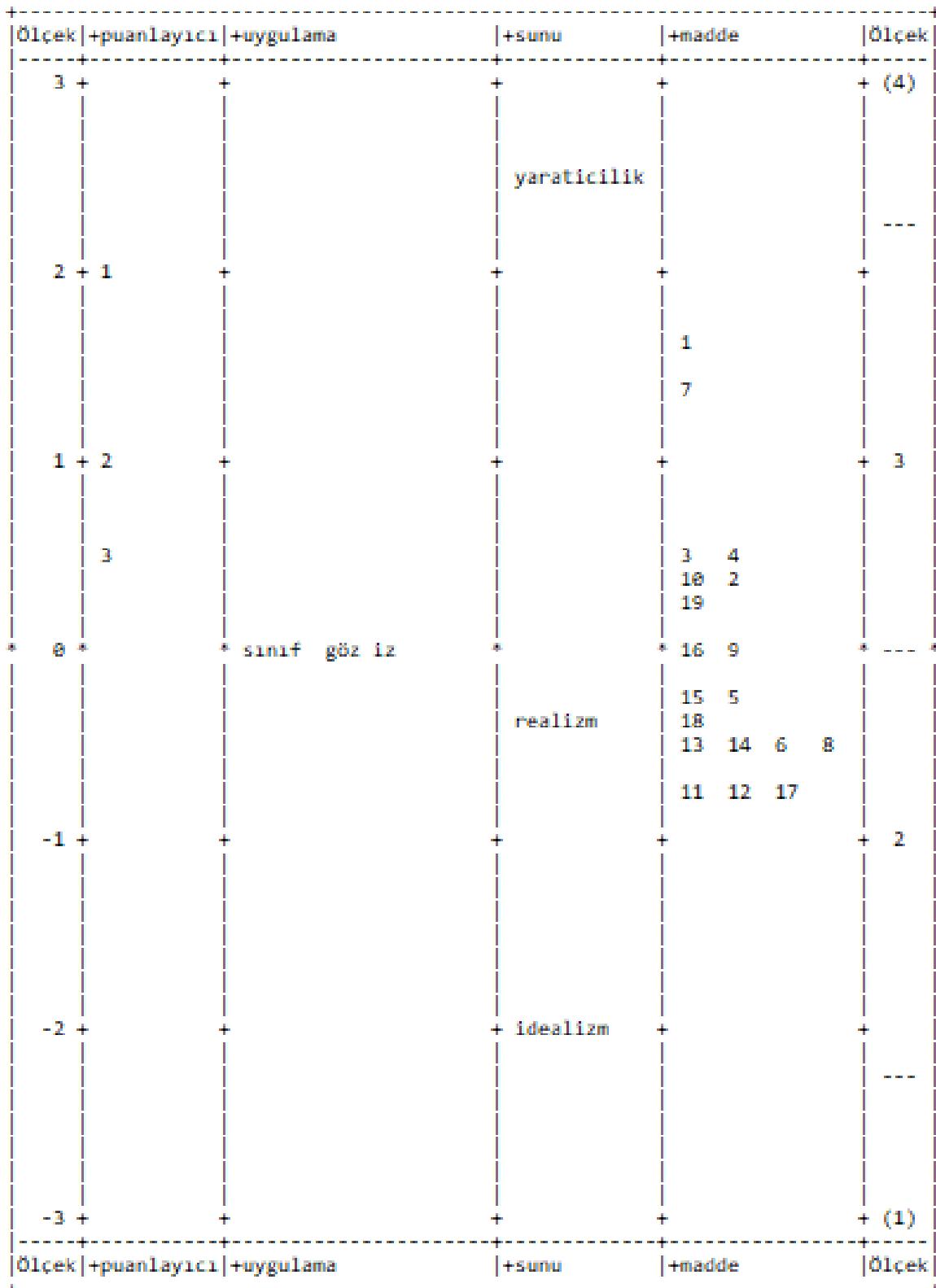
Üçüncü Alt Probleme İlişkin Bulgular

Araştırmanın üçüncü alt problemi “Performans değerlendirme sürecinde elde edilen puanlara uygulanan çok yüzeyli Rasch modeli sonuçları nasıldır?” şeklindedir. Bu alt problem doğrultusunda, sınıf içi ve göz izleme yöntemiyle ekran üzerinden verilen puanlar çok yüzeyli Rasch modeline göre analiz edilmiştir.

Çok yüzeyli Rasch analizi için Facet programı ile elde edilen değişken haritasında yüzeyler; üç farklı puanlayıcı, 2 farklı uygulama (sınıf içi ve tobi), üç farklı sunu (idealizm, realizm ve yaratıcılık) ve 19 maddeden oluşan sunum değerlendirme ölçeğinden oluşmaktadır. Çok yüzeyli Rasch analizi sonuçları Şekil 4’te görülmektedir.

Şekil 4

Çok Yüzeyle Rasch Analizi Değişken Haritası



Şekil 4'te gözlenen değişken haritasının ilk sütununda, logit ölçeği sonuçları sıralanmaktadır. Logit ölçeği, yüzeyler için elde edilen değerlerin konumlarını, -3 ile +3 aralığında, göstermektedir. Son sütunda ise sunum değerlendirme dereceli puanlama anahtarının derecelendirme sıralaması yer almaktadır. Değişken haritası incelendiğinde, genel olarak, aşağıdan yukarıya doğru madde güçlüğü'nün, puanlayıcı katılımının ve bireylerin yetenek düzeyinin artış gösterdiği gözlenmektedir (Doğru,2019; İlhan,2015; Wiseman, 2008).

Şekil 4'teki harita, puanlayıcıların, genellikle, katı bir puanlamaya sahip olduğunu göstermektedir. Puanlayıcılar tek tek ele alındığında; birinci puanlayıcının (P1) en düşük puanı verdiği ve en katı puanlamaya sahip olduğu, üçüncü puanlayıcının (P3) ise en yüksek puanı verdiği ve en cömert puanlamaya sahip olduğu söylenebilir. Sunumlar değerlendirildiğinde, en yüksek puanların genel olarak yaratıcılık sunumunda, en düşük puanların idealizm sunumunda verildiği bulunmuştur. Haritadan, sınıf içi ve göz izleme uygulamaları arasında bir uyum olduğu gözlenmiştir. Son olarak, madde güçlüklerinin yukarıdan aşağıya doğru azaldığı, en zor maddenin 1. madde ve en kolay maddenin 11.,12. ve 17. maddeler olduğu görülmektedir.

Facet programı ile sınıf içi ve göz izleme uygulamalarına ilişkin çok yüzeyli Rasch analiziyle elde edilen puanlayıcı yüzeyi ölçüm parametreleri Tablo 10'da verilmiştir.

Tablo 9

Sınıf İçi ve Göz İzleme Uygulaması Puanlarının Çok Yüzeyle Rasch Analiziyle Elde Edilen Puanlayıcı Yüzeyi Ölçüm Parametreleri

Puanlayıcı	Logit Ölçüsü	Standart Hata	İç Uyum (İnfit)	Dış Uyum (Outfit)
P1	1,94	0,16	0,65	0,51
P2	1,00	0,15	1,35	1,42
P3	0,44	0,15	0,93	1,02
Ortalama	1,13	0,16	0,98	0,98
Standart sapma (örneklem)	0,62	0,01	0,29	0,37
Standart sapma (evren)	0,76	0,01	0,35	0,46

Model, Evren: RMSE= 0,16, Standart Sapma=0,74, Ayırma Oranı=4,74, Güvenirlik=0,96

Model, Örneklem: RMSE= 0,16, Standart Sapma=0,60, Ayırma Oranı=3,82, Güvenirlik=0,95

Model, Tamamı Aynı Ki Kare=45,6, sd=2, p=0,00

Model, Rastgele Normal Ki Kare=1,9, sd=1, p=0,17

Puanlayıcılar arası mutlak uyum: %46,8

Puanlayıcılar arası beklenen uyum: %46,5

Puanlayıcılar arası güvenirlik Kappa İstatistiği: 0,01*

*Kappa İstatistiği=(Gözlenen Uyum-Beklenen Uyum)/(100-Beklenen Uyum) formülüyle hesaplanmıştır.

Tablo 9'da gözlenen ortalamalar, her bir puanlayıcının yapmış olduğu puanlamaların ortalamalarını belirtmektedir. Tablonun ilk sütununda yer alan logit değerleri, puanlayıcıların katılık/cömertlik açısından birbirleriyle karşılaştırılmasını sağlamaktadır. Logit değerinin pozitif olması değerlendirmenin katı olduğunu, negatif olması değerlendirmenin cömert olduğunu göstermektedir. Standart hata değerleri ise değerlendirmelerin doğruluğunu desteklemektedir. Tabloda yer alan iç uyum ve dış uyum değerlerinin 0,5 ve 1,5 arasında olması, puanlayıcıların yapmış olduğu değerlendirmelerinin uyumlu olduğunu göstermektedir (Linacre, 2018).

Üç puanlayıcıya ait facet sonuçlarının yer aldığı tablo incelendiğinde, logit değerlerinin ortalamasının 1,13 ve standart sapmasının 0,62 olduğu; puanlayıcılar için kestirilen iç uyum ve dış uyum değerlerin ortalamasının 0,98 olduğu gözlenmiştir. Bu değerler, 0,5 ve 1,5 arasında kabul edilebilir sınırlar içinde yer aldığı için model veri uyumunun olduğunu göstermektedir (Myford & Wolfe,2004).

Ayırma oranı ve güvenirlik indeksi, evren ve örneklem olmak üzere iki farklı modele göre hesaplanmaktadır. Linacre (2006), bir yüzeyin bütün bileşenlerinin model içerisinde yer alması durumunda model-evren, tüm bileşenlerinin yer almaması, tesadüfi olarak seçilmesi

durumunda ise model-örneklem sonuçlarının yorumlanmasını önermiştir. Model- örneklem sonuçlarına göre, ayırma oranı 3,82 ve güvenilirlik indeksi 0,95 olarak bulunmuştur. Puanlayıcılara ilişkin güvenilirlik indeksinin yüksek çıkması, puanlayıcılar arasında farklılık olduğu anlamına gelmektedir. Bu farklılığın anlamlılığı, ki kare değerine bakılarak yorumlanmaktadır (İlhan,2015), Her bir yüzey için tamamı aynı ki kare ve rastgele normal ki kare değeri elde edilmektedir. Yüzeydeki bileşenlerinin normal dağılım gösteren bir evrenden tesadüfi olarak seçilen bir örnekleme temsil etmesi durumunda, rastgele normal ki kare değeri kullanılmaktadır. Ölçme hatası göz önüne alınarak, yüzey bileşenleri arasında anlamlı fark olma durumunu belirlemek için tamamı aynı ki kare değeri incelenmektedir (İlhan, 2015; Linacre, 2018). Bu doğrultuda, tamamı aynı ki kare değeri anlamlı olduğu [$\chi^2 =45,6$, $sd=2$, $p<,01$] için puanlayıcılar arasında katılık ve cömertlik açısından anlamlı bir fark olduğu söylenebilir.

Tablo 9'a göre puanlayıcılar arası mutlak uyumun %46,8 ve puanlayıcılar arası beklenen uyumun %46,5 olduğu görülmektedir. Mutlak uyum, puanlayıcıların aynı sunuma ne derece aynı puanı verdiğini göstermektedir ve puanlayıcı güvenilirliğinin bir göstergesi olarak yorumlanabilmektedir. Yapılan çalışmalar mutlak uyumun en az %75 değerine ulaşması gerektiğini belirtmiştir (İlhan, 2015; Graham ve ark., 2012). Bu doğrultuda, puanlayıcılar arası uyumun düşük olduğu söylenebilir. Benzer şekilde, kappa istatistiğinin 0,01 olarak hesaplanması, puanlayıcılar arasında zayıf bir uyum olduğunu göstermektedir (Landis & Koch,1977).

Sınıf içi ve göz izleme uygulamalarına ilişkin çok yüzeyli Rasch analiziyle elde edilen uygulama yüzeyi ölçüm parametreleri Tablo 11'de verilmiştir.

Tablo 10

Sınıf İçi ve Göz İzleme Uygulaması Puanlarının Çok Yüzeyle Rasch Analiziyle Elde Edilen Uygulama Yüzeyi Ölçüm Parametreleri

Uygulama	Logit Ölçüsü	Standart Hata	İç Uyum	Dış Uyum
<i>Sınıf İçi</i>	0,05	0,13	1,17	1,20
<i>Tobi(Göz İzleme)</i>	-0,05	0,13	0,80	0,76
<i>Ortalama</i>	0,00	0,13	0,99	0,98
<i>Standart sapma (örneklem)</i>	0,05	0,00	0,19	0,22
<i>Standart sapma (evren)</i>	0,07	0,00	0,26	0,31

Model, Evren: RMSE= 0,13 Standart Sapma=0,00 Ayırma Oranı=0,00 Güvenirlik=0,00
 Model, Örneklem: RMSE= 0,13 Standart Sapma=0,00 Ayırma Oranı=0,00 Güvenirlik=0,00
 Model, Tamamı Aynı Ki Kare=1,5 sd=1 p=0,22

Göz izleme sonucu elde edilen logit değerleri puanlamaların cömert, sınıf içi uygulama sonucu elde edilen logit değerleri ise puanlamaların daha katı olduğunu göstermektedir. Uyum istatistiklerinin 0,5 ve 1,5 arasında olması, bu değerlerin kabul edilebilir aralıkta olduğu anlamına gelmektedir. Buna göre, sınıf içi ve göz izleme uygulaması sonucu elde edilen puanların uygun olduğu ve model veri uyumunun sağlandığı söylenebilir. Ayırma oranı ve güvenirlilik indeksi model-örneklem verileri dikkate alındığında, sınıf içi ve göz izleme uygulamalarında puanlayıcılar arası bir farklılık olmadığı bulunmuştur. Tabloda yer alan bir diğer istatistik olan tamamı aynı ki kare değeri incelendiğinde ise puanlama açısından uygulamalar arasında anlamlı bir farklılık bulunmadığı söylenebilir [$\chi^2=1,5$, $sd=1$, $p >0,01$].

Sınıf içi ve göz izleme uygulamalarına ilişkin çok yüzeyle Rasch analiziyle elde edilen sunu yüzeyi ölçüm parametreleri Tablo 11'de görülmektedir.

Tablo 11

Sınıf İçi ve Göz İzleme Uygulaması Puanlarının Çok Yüzeyle Rasch Analiziyle Elde Edilen Sunu Yüzeyi Ölçüm Parametreleri

Sunu	Logit Ölçüsü	Standart Hata	İç Uyum	Dış Uyum
<i>Yaratıcılık</i>	2,46	0,23	0,97	0,96
<i>Realizm</i>	-0,40	0,14	1,05	1,04
<i>İdealizm</i>	-2,06	0,14	0,93	0,95
<i>Ortalama</i>	0,00	0,17	0,98	0,98
<i>Standart sapma (örneklem)</i>	1,87	0,04	0,05	0,04
<i>Standart sapma (evren)</i>	2,29	0,05	0,06	0,05

Model, Evren: RMSE= 0,17 Standart Sapma=1,86 Ayırma Oranı=10,74 Güvenirlik=0,99
 Model, Örneklem: RMSE=0,17 Standart Sapma=2,28 Ayırma Oranı=13,17 Güvenirlik=0,99
 Model, Tamamı Aynı Ki Kare=294,1 sd=2 p=0,00
 Model, Rastgele Normal Ki Kare=2,00 sd=1 p=0,16

Tablo 11’de yer alan sunu yüzeyi ölçüm parametreleri, yaratıcılık sunusunun en katı puanlama değerlerine sahip olduğunu, realizm sunusunun puanlama katılık düzeyinin daha düşük olduğunu ve idealizm sunusunun en cömert puanlama düzeyine sahip olduğunu göstermektedir. Hesaplanan iç ve dış uyum istatistikleri kabul edilebilir değerler arasında yer almaktadır. Model-örneklem ayırma oranı ve güvenirlik indeksi değerleri incelendiğinde, uygulama konusuna göre puanlamaların anlamlı bir farklılık gösterdiği söylenebilir. Tamamı aynı ki kare değeri incelendiğinde, sunu konusu değişikçe puanlamalar arasında anlamlı bir farklılığın olduğu görülmektedir [$\chi^2=294,1$, $sd=2$, $p < 0,01$].

Sınıf içi ve göz izleme uygulamalarına ilişkin çok yüzeyle Rasch analiziyle elde edilen madde yüzeyi ölçüm parametreleri Tablo 12’de yer almaktadır.

Tablo 12

Sınıf İçi ve Göz İzleme Uygulaması Puanlarının Çok Yüzeyle Rasch Analiziyle Elde Edilen Madde Yüzeyi Ölçüm Parametreleri

Madde	Logit Ölçüsü	Standart Hata	İç Uyum	Dış Uyum
1	1,63	0,45	0,46*	0,97
7	1,43	0,44	0,56	0,50
3	0,56	0,40	0,90	0,85
4	0,56	0,40	1,76*	1,29
2	0,40	0,40	1,14	1,34
10	0,40	0,40	1,40	1,46
19	0,25	0,39	1,42	2,16*
9	-0,06	0,39	0,58	0,62
16	-0,06	0,39	0,61	1,52*
5	-0,21	0,38	1,05	0,90
15	-0,21	0,38	1,79*	1,52*
18	-0,35	0,38	0,55	0,58
6	-0,50	0,38	0,88	0,79
8	-0,50	0,38	0,57	0,59
13	-0,50	0,38	0,51	0,50
14	-0,50	0,38	1,37	1,16
11	-0,79	0,38	1,28	1,22
12	-0,79	0,38	0,97	0,90
17	-0,79	0,38	0,81	0,77
<i>Ortalama</i>	0,00	0,39	0,98	0,98
<i>Standart sapma (örneklem)</i>	0,68	0,02	0,41	0,43
<i>Standart sapma (evren)</i>	0,70	0,02	0,43	0,44
Model, Evren: RMSE=0,39 Standart Sapma=0,55 Ayırma Oranı=1,41 Güvenirlik=0,66				
Model, Örneklem: RMSE=0,39 Standart Sapma=0,58 Ayırma Oranı=1,46 Güvenirlik=0,68				
Model, Tamamı Aynı Ki Kare=50,9 sd=18 p=0,00				
Model, Rastgele Normal Ki Kare=13,8 sd=17 p=0,68				

Tablo 12 incelendiğinde, maddelerin 1,63 ile -0,79 arasında değişen logit değerlerine sahip olduğu görülmektedir. İç uyum ve dış uyum istatistikleri, bazı maddelerin (1.,4.,15.,16. ve 19. maddeler) kabul edilebilir değerler arasında yer almadığını göstermektedir. Bu maddelerin öğrencilerin sunum yapma performanslarını tutarsız puanladığı şeklinde yorumlanmaktadır. Tablo 12 incelendiğinde, güvenirlilik indeksi ve ayırma oranının kabul edilebilir düzeyde olduğu söylenebilir. Güvenirlilik indeksinin yüksek olması, madde yüzeyi için kavramsal olarak farklı maddelerin birbirinden ayırt edilebilir olduğu anlamına gelmektedir.

Tamamı aynı ki kare değeri ise maddeler değişikçe puanlamalar arasında anlamlı bir farklılığın olduğuna işaret etmektedir [$\chi^2=50,9$, $sd=18$, $p <0,01$].

Sınıf içi ve göz izleme uygulamalarına ilişkin çok yüzeyli Rasch analiziyle elde edilen kategori istatistikleri Tablo 13'te verilmiştir.

Tablo 13

Puanlama Sonuçlarının Çok Yüzeyli Rasch Modeline Göre Kategori Analiz Sonucu İstatistikleri

Ölçek Kategorisi	Frekans	Yüzde	Yığılmalı Yüzde	Ortalama Ölçüm	Dış Uyum
1	34	10	10	-1,21	1,20
2	85	25	35	-0,66	0,90
3	104	30	65	1,03	0,90
4	119	35	100	3,15	1,10

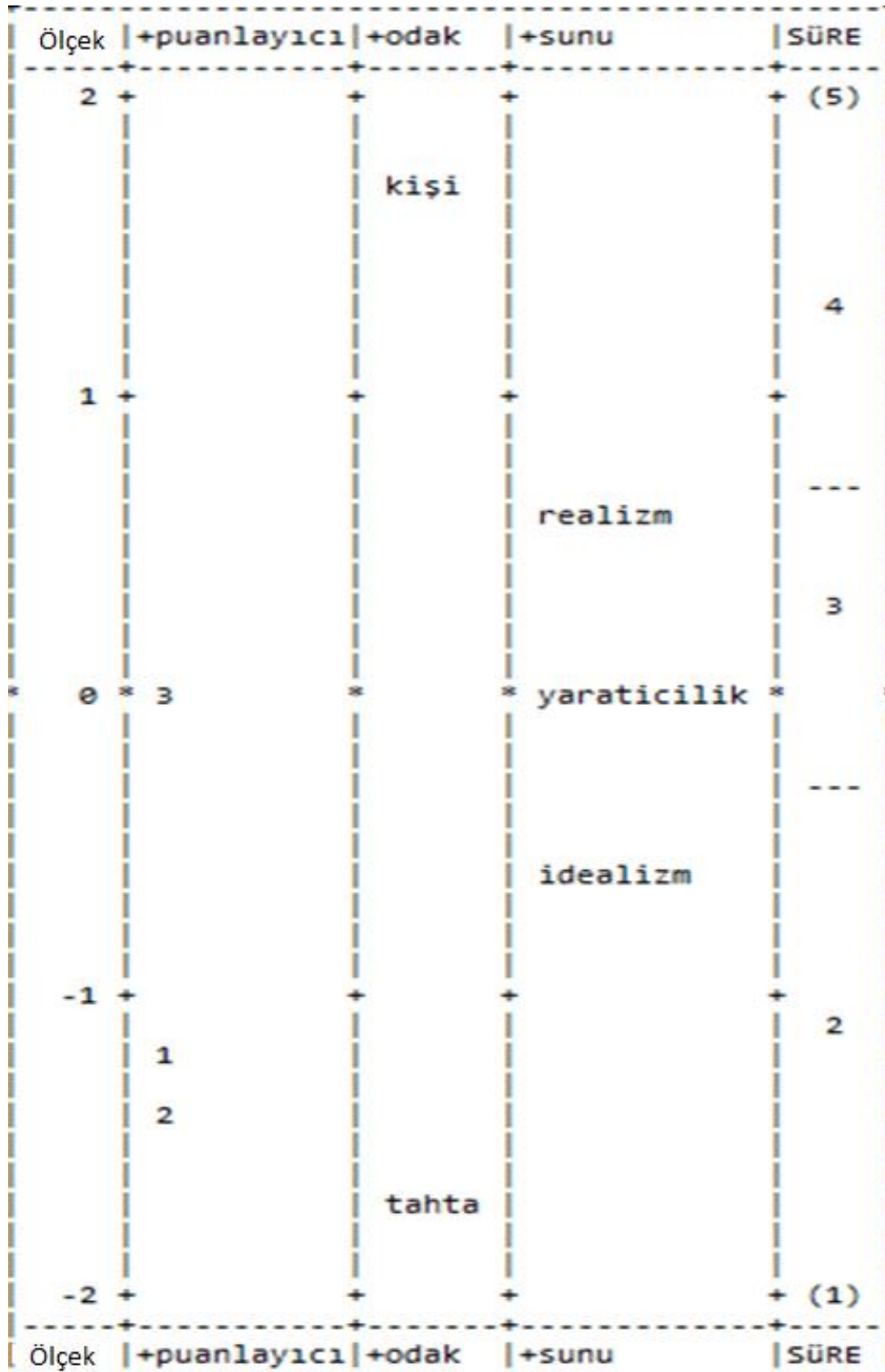
Tablo 13'ün birinci sütununda, sunum değerlendirme dereceli puanlama anahtarı kategori sayısı yer almaktadır (tamamen yeterli: 4, yeterli: 3, kısmen yeterli: 2 ve yeterli değil: 1). Her bir kategori değeri için frekans, yüzde, yığılmalı yüzde, ortalama ölçüm ve dış uyum değerleri hesaplanmıştır. İlhan (2015), her bir kategorinin frekans değerlerin en az 10 olması gerektiğini belirtmiştir. Ortalama ölçümlerin monoton artması ve uygunluk dışı değerlerin ikiden düşük değerler alması puanlama ölçeğinin uygunluğunu göstermektedir (Linacre, 2018).

Dördüncü Alt Probleme İlişkin Bulgular

Araştırmanın dördüncü alt problemi "Göz izleme sonucu puanlayıcılardan elde edilen göz izleme metriklerine uygulanan çok yüzeyli Rasch modeli sonuçları nasıldır?" şeklindedir. Bu alt problem doğrultusunda, puanlayıcıların ortalama sabitlenme süresi metriklerinden elde edilen veriler çok yüzeyli Rasch modeline göre analiz edilmiş ve analiz sonuçları Şekil 5'te sunulmuştur.

Şekil 5

Göz İzleme Tekniğiyle Elde Edilen Ortalama Sabitleme Metriklerinin Çok Yüzeysel Rasch Analizi Değişken Haritası



Şekil 5'te gözlenen ortalama sabitleme süreleri incelendiğinde, puanlayıcıların en çok sunumu yapan kişiye odaklandığı görülmektedir.

Tablo 14'te göz izleme uygulaması ortalama sabitleme süreleri metrikleri için çok yüzeyli Rasch analiziyle elde edilen puanlayıcı yüzeyi ölçüm parametreleri yer almaktadır.

Tablo 14

Ortalama Sabitleme Süreleri Metrikleri Çok Yüzeyli Rasch Analiziyle Elde Edilen Puanlayıcı Yüzeyi Ölçüm Parametreleri

Puanlayıcı	Logit Ölçüsü	Standart Hata	İç Uyum	Dış Uyum
P3	-0,03	0,59	0,36	0,48
P1	-1,16	0,52	0,44	0,77
P2	-1,44	0,53	0,37	0,98
<i>Ortalama</i>	-0,88	0,55	0,72	0,74
<i>Standart sapma (örneklem)</i>	0,61	0,03	0,46	0,21
<i>Standart sapma (evren)</i>	0,74	0,03	0,56	0,25

Model, Evren: RMSE= 0,55 Standart Sapma=0,26 Ayırma Oranı=0,48 Güvenirlik=0,19
 Model, Evren: RMSE= 0,55 Standart Sapma=0,51 Ayırma Oranı=0,92 Güvenirlik=0,46
 Model, Tamamı Aynı Ki Kare=3,5 sd=2 p=0,17
 Model, Rastgele Normal Ki Kare=1,4 sd=1 p=0,23
 Puanlayıcılar arası mutlak uyum: %33,3
 Puanlayıcılar arası beklenen uyum: %36,4

Tablo 14'teki puanlayıcı yüzeyi ölçüm parametreleri incelendiğinde, P1, P2 ve P3 iç uyum değerleri kabul edilebilir sınırların dışında olduğu belirlenmiştir. Bu durum, ortalama sabitleme süreleri açısından P1, P2 ve P3'ün ortalama sabitleme sürelerine göre farklılık gösterdiğine işaret etmektedir. Güvenirlik indeksi incelendiğinde, puanlayıcıların ortalama sabitleme sürelerine göre anlamlı bir farklılık göstermediği söylenebilir. Tamamı aynı ki kare değeri incelendiğinde, puanlayıcılar arasında ortalama sabitleme süreleri açısından anlamlı bir farkın olmadığı bulunmuştur [$\chi^2=3,5$, $sd=2$, $p > 0,01$]. Puanlayıcılar arası mutlak uyumun %33,3 ve beklenen uyumun %36,4 olması puanlayıcılar arası güvenirliliğin düşük olduğunu göstermektedir.

Tablo 15'de göz izleme uygulaması ortalama sabitleme süreleri metrikleri çok yüzeyli Rasch analiziyle elde edilen odak noktası yüzeyi ölçüm parametreleri yer almaktadır.

Tablo 15

Ortalama sabitleme Süreleri Metrikleri Çok Yüzeyle Rasch Analiziyle Elde Edilen Odak Noktası Yüzeyle Ölçüm Parametreleri

Uygulama	Logit Ölçüsü	Standart Hata	İç Uyum	Dış Uyum
<i>Kişi</i>	1,75	0,36	0,73	0,65
<i>Tahta</i>	-1,75	0,62	0,79	0,84
<i>Ortalama</i>	0,00	0,49	0,76	0,74
<i>Standart sapma (örneklem)</i>	1,75	0,13	0,03	0,10
<i>Standart sapma (evren)</i>	2,47	0,18	0,04	0,14

Model, Evren: RMSE= 0,51 Standart Sapma=1,67 Ayırma Oranı=3,27 Güvenirlik=0,91
 Model, Örneklem: RMSE= 0,51 Standart Sapma=2,42 Ayırma Oranı=4,74 Güvenirlik=0,96
 Model, Tamamı Aynı Ki Kare=23,4 sd=1 p=0,00

Tablo 15’de logit değerleri incelendiğinde, ortalama sabitleme süresine göre kişinin tahtaya göre daha fazla ilgi odağı olduğu görülmektedir. Hesaplanan iç ve dış uyum istatistiklerinin kabul edilebilir sınırlar içerisinde olması, puanlayıcıların ilgi odakları açısından ortalama sabitleme süreleri için oluşturulan modelin veri uyumunu sağladığını göstermektedir. Ayırma oranı ve güvenirlilik indeksi değerleri, bireylerin ilgi odakları açısından yaptıkları değerlendirmelerinin birbirinden anlamlı olarak farklılaştığına işaret etmektedir. Tamamı aynı ki kare değerine göre ilgi odakları açısından ortalama sabitleme sürelerinin puanlayıcılar arasında anlamlı olarak farklılaştığı bulunmuştur [$\chi^2=23,4$, $sd=1$, $p < 0,01$].

Tablo 16’da göz izleme uygulaması ortalama sabitleme süreleri metrikleri çok yüzeyle Rasch analiziyle elde edilen sunu yüzeyle ölçüm parametreleri yer almaktadır.

Tablo 16

Ortalama sabitleme Süreleri Metrikleri Çok Yüzeyle Rasch Analiziyle Elde Edilen Sunu Yüzeyi Ölçüm Parametreleri

Sunu	Logit Ölçüsü	Standart Hata	İç Uyum	Dış Uyum
<i>Yaratıcılık</i>	0,00	0,53	1,36	1,09
<i>Realizm</i>	0,56	0,55	0,31	0,39
<i>İdealizm</i>	-0,56	0,55	0,51	0,76
<i>Ortalama</i>	0,00	0,55	0,73	0,74
<i>Standart sapma (örneklem)</i>	0,46	0,01	0,45	0,29
<i>Standart sapma (evren)</i>	0,56	0,01	0,56	0,35

Model, Evren: RMSE= 0,55 Standart Sapma=0,00 Ayırma Oranı=0,00 Güvenirlik=0,00
 Model, Örneklem: RMSE=0,55 Standart Sapma=0,13 Ayırma Oranı=0,23 Güvenirlik=0,06
 Model, Tamamı Aynı Ki Kare=294,1 sd=2 p=0,00
 Model, Rastgele Normal Ki Kare=2,00 sd=1 p=0,16

Tablo 16'da göz izleme uygulaması ortalama sabitleme süreleri metrikleri çok yüzeyle Rasch Analiziyle elde edilen sunu yüzeyi ölçüm parametreleri incelendiğinde yaratıcılık ve idealizm sunu yüzeyinde model veri uyumuna uygun ortalama sabitleme sürelerine sahip oldukları yorumu yapılabilir. Uyum istatistiklerinin 0,5 ve 1,5 arasında olması, bu değerlerin kabul edilebilir aralıkta olduğu anlamına gelmektedir. Realizm sunumunda iç uyum değerinin kabul edilen sınırlar dışında olduğu belirlenmiştir. Bu durum realizm sunumunda ortalama sabitleme sürelerinin yaratıcılık ve idealizm sunumuna göre farklılaştığı şeklinde yorumlanmaktadır. Ayırma oranı ve güvenirlik indeksi incelendiğinde sunuya göre ortalama sabitleme süreleri arasında anlamlı bir farklılık bulunmamaktadır. Tamamı aynı ki kare [$\chi^2=294$, $sd=2$, $p >0,01$] değeri incelendiğinde yine anlamlı bir farklılık bulunmadığı yorumu yapılabilir.

Tablo 17'de Göz izleme uygulaması ortalama sabitleme süreleri metrikleri çok yüzeyle Rasch Analiziyle elde edilen Kategori İstatistikleri yer almaktadır.

Tablo 17

Göz İzleme Uygulaması Ortalama Sabitleme Süreleri Metrikleri Çok Yüzeysel Rasch Analiziyle Elde Edilen Kategori İstatistikleri

Ölçek Kategorisi	Frekans	Yüzde	Yığılmalı Yüzde	Ortalama Ölçüm	Dış Uyum
1	5	28	28	-2,85	0,9
2	6	33	61	-1,60	0,7
3	2	11	72	0,73	0,2
4	3	17	89	1,34	0,1
5	2	11	100	1,29	1,4

Tablo 17 'de kategori istatistikleri yer almaktadır. Birinci sütunda ortalama sabitleme sürelerinin kategori sayısı yer almaktadır. İkinci sütunda ise her bir kategorinin kaç defa kullanıldığı yer almaktadır. Her bir kategori değeri için frekans, yüzde, yığılmalı yüzde değerleri Tablo 17'de yer almaktadır. Yapılan çalışmalarda her bir kategorinin en az on değerinde bulunması ve ortalama ölçümlerin monoton artışı ve uyum dışı değerlerin 0,5 ile 1,5 arasında değerler alması puanlama ölçeğinin uygunluğunu göstermektedir (Linacre, 2018).

Beşinci Alt Probleme İlişkin Bulgular

Araştırmanın beşinci alt problemi "Göz izleme sonucunda puanlayıcılardan elde edilen göz izleme metrikleri nasıldır?" şeklinde ifade edilmiştir. Göz izleme uygulamasından elde edilen metriklerle ait tablolar ile yol ve sıcaklık haritaları İdealizm, Yaratıcılık ve Realizm sunumları için ayrı ayrı incelenmiş ve sonuçlar yorumlanmıştır.

Tablo 18'de Göz izleme uygulaması sonucunda idealizm sunumunda toplam sabitleme sürelerine ait betimsel istatistikler yer almaktadır.

Tablo 18*İdealizm Sunumunda Toplam Sabitlenme Sürelerine(Sn) Ait Betimsel İstatistikler*

İlgi Alanı Üzerinde Toplam Sabitlenme Süresi	İLt1	İLö1	Ortalama	Medyan	Toplam	Toplam Sabitlenme Süresi	Toplam Kayıt Süresi
<i>P1</i>	217,9	395,8	306,8	306,8	613,7	786,9	1456,7
<i>P2</i>	262,6	200,4	231,5	231,5	463,0	750,8	869,3
<i>P3</i>	313,0	246,8	279,9	279,9	559,8	748,2	955,1
<i>Ortalama</i>	264,5	281,0	272,8	272,8	545,5	762,0	1307,5
<i>Toplam süredeki oranı(%)</i>	48,5	51,5					
<i>Varyans</i>	2262,6	10420,3	1458,2	1458,2	5832,8	468,0	1393886,4
<i>Standart sapma</i>	47,6	102,1	38,2	38,2	76,4	21,6	1180,6

İLt1: ilgi alanı tahta idealizm; İLö1: İlgi alanı öğrenci idealizm

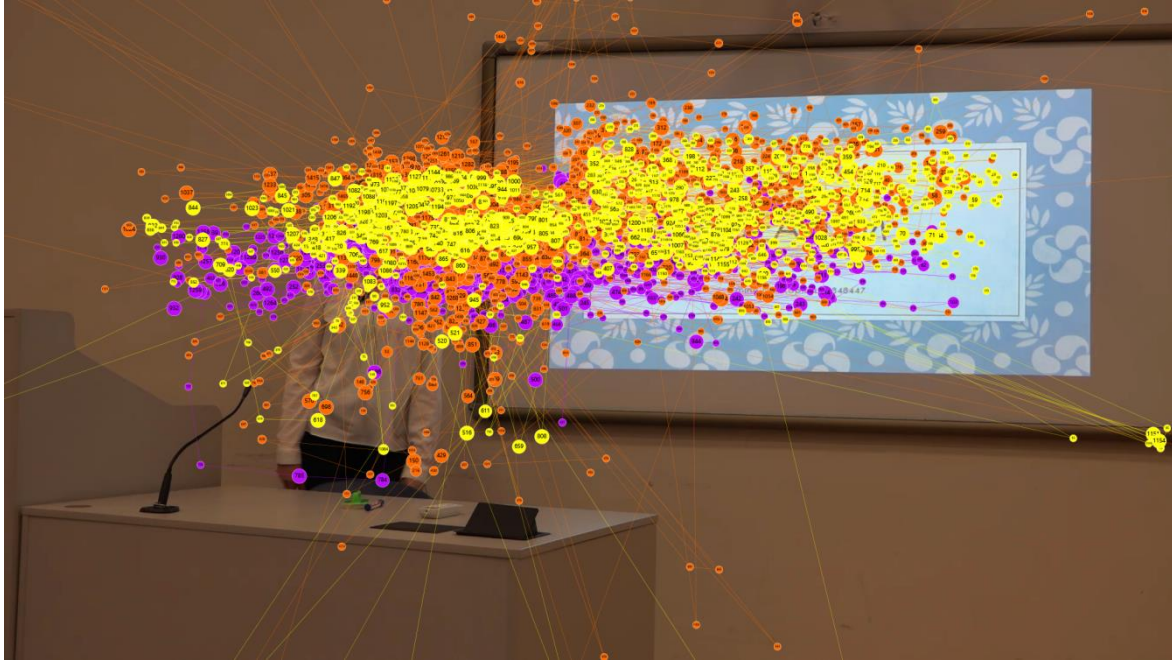
Tablo 18 incelendiğinde, toplam kayıt süreleri $P1 > P3 > P2$ şeklinde sıralanmaktadır. Toplam sabitlenme süreleri ise $P1 > P2 > P3$ şeklinde sıralanmaktadır. İlgi alanı üzerinde toplam sabitlenme süreleri ise $P1 > P3 > P2$ şeklinde sıralanmaktadır. Toplam sabitlenme süresi ile ilgi alanı üzerinde toplam sabitlenme süresi arasındaki fark artıka Puanlayıcıların dikkatinin dağıldığı yorumu yapılabilmektedir (Jian,2016). Buna göre farkın en fazla olduğu P2 dikkati en fazla dağılan puanlayıcı, P1 ise dikkati en az dağılan puanlayıcıdır.

Şekil 1 de belirtilen iki ilgi alanı üzerinden analiz yapılmaktadır (Sunumu yapan birey ve sunum yapılan tahta). İlgi alanı üzerinde toplam sabitlenme süresine göre puanlayıcıların göz sabitlenmelerinin %51'i sunum yapan kişi üzerinde %48'i ise sunum yapılan tahta üzerinde gerçekleşmiştir. P1'in ilgi alanları üzerindeki sabitlenme süresi incelendiğinde sunum yapan birey üzerinde, P2 ve P3 sunum yapılan tahta üzerinde daha çok zaman geçirmişler. İlgi alanı üzerinde toplam sabitlenme süreleri görsel Şekil 6'da yol haritasında ve Şekil 7 'de ısı haritasında da incelenebilmektedir. Toplam kayıt süreleri farklı olsa bile puanlayıcıların ilgi alanı üzerinde geçirdikleri zamanlar birbirlerine oldukça yakın değerler almıştır.

Şekil 6' da İdealizm sunumuna ait göz izleme metrikleri ile elde edilen yol haritası yer almaktadır.

Şekil 6

İdealizm Sunumu Yol Haritası



P1: sarı; P2: turuncu; P3: mor renk ile kodlanmıştır.

Şekil 6 'da yer alan İdealizm Sunumuna ait yol haritası incelendiğinde puanlayıcıların farklı renkte olan yol haritalarının birbirine benzer olduğu gözlemlenmektedir. Yol haritasında sabitleme noktalarının rotası ve sabitleme noktasındaki süresine göre dairelerin alanı büyüyüp küçülmektedir. Dairelerin içerisinde ise Puanlayıcıların sabitleme sıraları yer almaktadır. Birinci puanlayıcı sarı, ikinci puanlayıcı turuncu ve üçüncü puanlayıcı mor renk ile işaretlenmiş durumdadır. Şekil 6'da yol haritası incelendiğinde puanlayıcıların ilgi alanları dışına odaklanmadığı, sunum performansını izlerken dikkatlerinin dağılmadığı yorumunu yapabiliriz (Jian,2017).

Şekil 7' de İdealizm sunumuna ait göz izleme metrikleri ile elde edilen sıcaklık haritası yer almaktadır.

Şekil 7*İdealizm Sunumu Sıcaklık Haritası*

Şekil 7'de yer alan idealizm sunumu için elde edilen sıcaklık haritası incelendiğinde sunum yapan öğrenci ilgi alanı (İLÖ1) üzerinde yoğunlaşmanın daha fazla olduğu gözlemlenmektedir. Sunum yapılan tahta üzerinde ise (İLt1) yoğunlaşmanın az olduğu gözlemlenmektedir. Sıcaklık haritasında yoğunluğun en fazla olduğu alan kırmızı daha sonra sarı en az yoğunluk yeşil renk ile ifade edilmektedir.

Tablo 19'da Göz izleme uygulaması sonucunda yaratıcılık sunumunda toplam sabitleme sürelerine ait betimsel istatistikler yer almaktadır.

Tablo 19*Yaratıcılık Sunumu Toplam Sabitleme Sürelerine Ait Betimsel İstatistikler*

Toplam Sabitleme Süresi	İlt2	İlö2	Ort	Med	Top	Toplam Sabitleme Süresi	Toplam Kayıt Süresi
<i>P1</i>	256,3	1242,9	749,6	749,6	1499,2	2924,9	3137,7
<i>P2</i>	490,2	1445,6	967,9	967,9	1935,8	2933,9	3002,5
<i>P3</i>	231,6	1645,3	938,5	938,5	1876,9	2942,3	3132,0
<i>Ort</i>	326,1	1444,6	885,3	885,3	1770,7	2933,7	1573,4
<i>Toplam süredeki oranı (%)</i>	18,4	81,6					
<i>Varyans</i>	20354,6	40476,3	14028,8	14028,8	56115,3	75,3	1393886,4
<i>Standart sapma</i>	142,7	201,2	118,4	118,4	236,9	8,7	1180,6

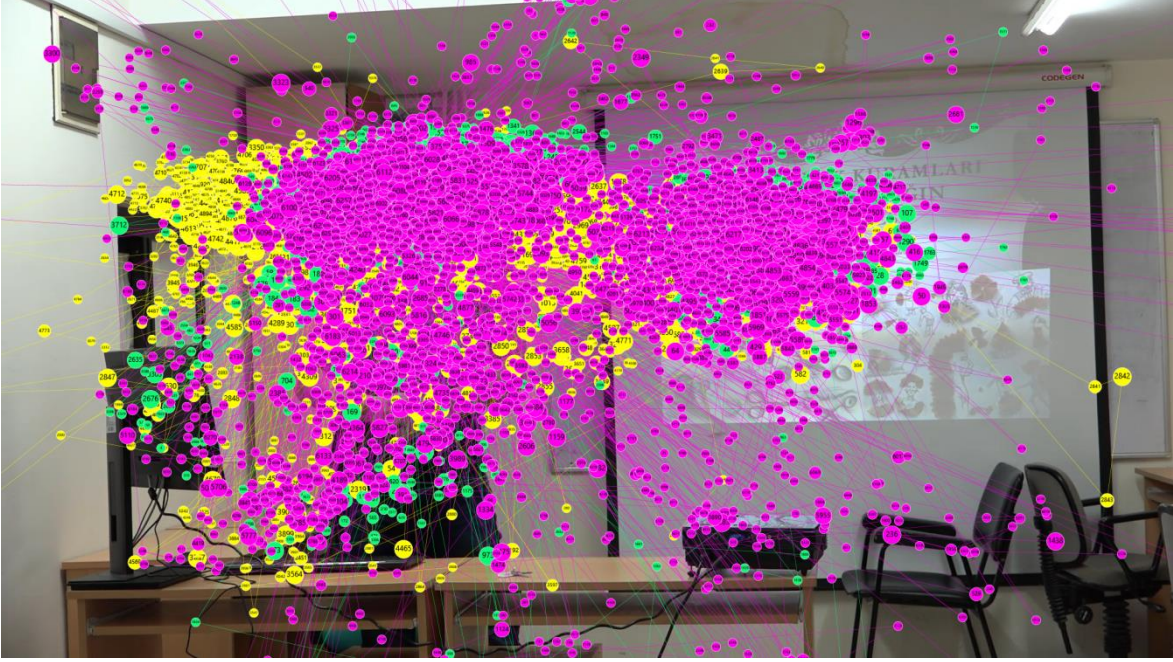
İlt2: ilgi alanı tahta yaratıcılık; İlö2: ilgi alanı öğrenci yaratıcılık

Tablo 19'da yer alan İLÖ2 ve İLT2 ilgi alanları karşılaştırılması sonucunda, Puanlayıcıların toplam sabitleme süreleri %81'i sunum yapan kişi üzerinde %18'i ise sunum yapılan tahta üzerinde gerçekleşmiştir. Toplam kayıt süreleri $P1 > P3 > P2$ şeklinde sıralanmaktadır. Toplam sabitleme süreleri ise $P3 > P2 > P1$ şeklinde tam tersi olarak sıralanmaktadır. İlgi alanı üzerinde toplam sabitleme ise $P2 > P3 > P1$ şeklinde sıralanmaktadır. Toplam sabitleme süresi ile ilgi alanı üzerindeki toplam sabitleme süre arasındaki fark artıkça puanlayıcının odaklanma sürecinde olumsuzluk yaşadığı ve dikkatinin dağıldığı yorumu yapılabilmektedir (Jian,2017). P2 dikkati en az dağılan P1 ise dikkati en fazla dağılan puanlayıcıdır şeklinde yorumlayabiliriz.

Şekil 8' de yaratıcılık sunumuna ait göz izleme metrikleri ile elde edilen yol haritası yer almaktadır.

Şekil 8

Yaratıcılık Sunumu Yol Haritası



P1:yeşil; P2:sarı;P3:mor renk ile kodlanmıştır.

Şekil 8'de yer alan yaratıcılık sunumuna ait yol haritası incelendiğinde puanlayıcıların farklı renkte olan yol haritalarının birbirine benzer olduğu gözlemlenmektedir. Birinci puanlayıcı yeşil, ikinci puanlayıcı sarı ve üçüncü puanlayıcı mor renk ile işaretlenmiş durumdadır. Puanlayıcıların yaratıcılık sunumunda daha fazla dikkatlerinin dağıldığını dairelerin çoğalması ve ilgi alanı dışına çıkması nedeniyle söyleyebiliriz.

Şekil 9' da yaratıcılık sunumuna ait göz izleme metrikleri ile elde edilen sıcaklık haritası yer almaktadır.

Şekil 9*Yaratıcılık Sunumu Sıcaklık Haritası*

Şekil 9'da yer alan yaratıcılık sunumu için elde edilen sıcaklık haritası incelendiğinde sunum yapan öğrenci ilgi alanı (İLÖ2) üzerinde yoğunlaşmanın daha fazla olduğu gözlemlenmektedir. Sunum yapılan tahta üzerinde ise (İLt2) yoğunlaşmanın az olduğu gözlemlenmektedir.

Tablo 20'de Göz izleme uygulaması sonucunda Realizm sunumunda toplam sabitleme sürelerine ait betimsel istatistikler yer almaktadır.

Tablo 20*Realizm Sunumu Toplam Sabitlenme Sürelerine Ait Betimsel İstatistikler*

Toplam Süresi	Sabitlenme	İLt3	İLö3	Ort	Med	Top	Toplam Sabitlenme Süresi	Toplam Kayıt Süresi
P1		33,45	196,23	114,84	114,84	229,68	273,56	808,02
P2		7,71	215,85	111,78	111,78	223,56	289,28	395,34
P3		63,36	144,28	103,82	103,82	207,64	270,69	404,10
ort		34,84	185,45	110,15	110,15	220,29	277,84	1573,42
Toplam süredeki oran (%)		15,82	84,18					
Varyans		775,84	1367,56	32,35	32,35	129,39	100,13	1393886,43
Standart sapma		27,85	36,98	5,69	5,69	11,37	10,01	1180,63

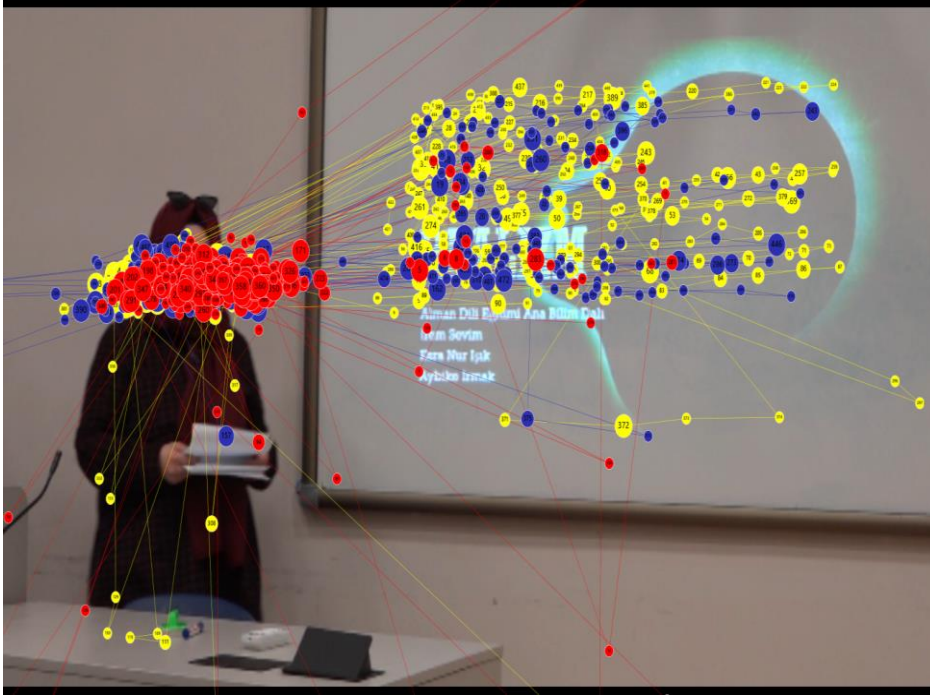
İLt3: ilgi alanı tahta realizm; İLö3: İlgi alanı öğrenci realizm

Tablo 20 incelendiğinde puanlayıcıların sabitlenme süresinin % 84'ünün sunum yapan kişiye (İLö3), % 15'inin sunum yapılan tahtaya (İLt3) yönelik olduğu görülmektedir. Toplam kayıt sürelerine göre P1 > P3 > P2 şeklinde bir sıralama elde edilmektedir. Toplam Sabitlenme süreleri ise P2 > P1 > P3 şeklinde sıralanmaktadır. İlgi alanı üzerinde toplam sabitlenme süreleri ise P1 > P2 > P3 şeklinde sıralanmaktadır. İlgi alanı üzerinde toplam sabitlenme süresi ile toplam sabitlenme süresi arasındaki fark en fazla P3 en az ise P1 'in değerlerinde gözlemlenmektedir. P3 en fazla dikkati dağılan P1 ise en az dikkati dağılan puanlayıcıdır şeklinde ifade edilebilir.

Şekil 10' da realizm sunumuna ait göz izleme metrikleri ile elde edilen yol haritası yer almaktadır.

Şekil 10

Realizm Sunumu Yol Haritası



P1:mavi; P2: kırmızı; P3: sarı renkle kodlanmıştır.

Şekil 10'de yer alan Realizm Sunumuna ait yol haritası incelendiğinde puanlayıcıların farklı renkte olan yol haritalarının birbirine benzer olduğu gözlemlenmektedir. Birinci puanlayıcı mavi, ikinci puanlayıcı kırmızı ve üçüncü puanlayıcı sarı renk ile işaretlenmiş durumdadır. Genel olarak bütün puanlayıcıların yol haritası içerisinde ilgi alanları dışına çok nadir odaklandığı gözlemlenmektedir. Buradan puanlayıcıların dikkatinin dağılmadığı ve puanlama yaparken odaklandıkları yorumu yapılabilir.

Şekil 11' de realizm sunumuna ait göz izleme metrikleri ile elde edilen sıcaklık haritası yer almaktadır.

Şekil 11

Realizm Sunumu Sıcaklık Haritası



Şekil 11’de yer alan Realizm Sunumu için elde edilen sıcaklık haritası incelendiğinde sunum yapan öğrenci ilgi alanı (İLÖ3) üzerinde yoğunlaşmanın daha fazla olduğu gözlemlenmektedir. Sunum yapılan tahta üzerinde ise (İLt3) yoğunlaşmanın az olduğu gözlemlenmektedir. Genel olarak bütün performans görevleri için sıcaklık haritaları incelendiğinde bütün puanlayıcıların belirlenen ilgi alanları dışına çok fazla odaklanmadıkları ve özellikle sunum yapan öğrencilere daha fazla odaklandıkları gözlemlenmiştir. Bunun yanında süresi daha fazla olan yaratıcılık sunumunda puanlayıcıların dikkatinin biraz daha fazla dağıldığı ve zaman zaman ilgi alanı dışına göz hareketlerinin çıktığı gözlenmektedir.

Puanlayıcılar, göz izleme yöntemiyle elde edilen metrikleri; ortalama sabitlenme süresi, sabitlenme sayısı, ilk sabitlenmeye kadar geçen süre, toplam kayıt süresi, ilk sabitlenme süresi, ilgi alanı üzerinde toplam sabitlenme süresidir. Her bir performans görevinin süresi farklı olduğu için, puanlayıcıların toplam sabitlenme sürelerinin karşılaştırılması anlamlı olmayacaktır. Bu nedenle ortalama sabitlenme sürelerine çok yüzeysel Rasch analizi uygulanmıştır. Toplam sabitlenme süreleri ise puanlayıcıların, belirlenen ilgi alanlarına göre betimsel istatistikleri ve bu istatistiklerin görsel sunumunu veren yol ve sıcaklık haritaları ile birlikte yorumlanmıştır. Bu aşamadan sonra göz izleme yöntemiyle elde edilen metriklerin birlikte daha anlaşılır yorumlanabilmesi için genel bir göz izleme metrikleri tablosu oluşturuldu.

Tablo 22’de Göz izleme uygulaması ile elde edilen puanlayıcılara ait metriklerin genel tablosu yer almaktadır.

Tablo 21

Göz İzleme Uygulaması ile Elde Edilen Puanlayıcılara Ait Metrikler Genel Tablosu

		İlgi Alanı Üzerinde Toplam Sabitlenme Süresi	Ortalama Sabitlenme Süresi	Sabitlenme Sayısı	İlk Sabitlenmeye Kadar Geçen Süre	İlk Sabitlenme Süresi
<i>Realizm sunum yapan</i>	P1	196,23	0,55	356,00	0,00	0,54
	P2	144,28	0,71	303,00	0,00	0,02
	P3	215,85	0,63	230,00	0,00	0,24
<i>Realizm tahta</i>	P1	33,45	0,26	128,00	7,89	0,57
	P2	63,36	0,27	29,00	0,73	0,15
	P3	7,71	0,29	216,00	2,88	0,32
<i>İdealizm sunum yapan</i>	P1	395,77	0,63	629,00	0,00	0,25
	P2	200,37	0,53	379,00	0,00	1,26
	P3	246,85	0,78	317,00	0,13	0,78
<i>İdealizm tahta</i>	P1	217,92	0,42	518,00	0,28	0,10
	P2	262,61	0,31	839,00	2,52	0,37
	P3	313,00	0,42	742,00	1,17	1,22
<i>Yaratıcılık sunum yapan</i>	P1	1242,91	0,45	2736,00	0,00	0,20
	P2	1445,62	0,40	3624,00	0,23	0,23
	P3	1645,28	0,63	2621,00	0,00	0,07
<i>Yaratıcılık tahta</i>	P1	256,34	0,39	657,00	8,99	0,23
	P2	490,18	0,30	1635,00	3,74	0,16
	P3	231,65	0,41	560,00	4,89	0,20

Tablo 21 ‘de göz izleme tekniğiyle elde edilen puanlayıcılara ait metrikler; ilk sabitlenmeye kadar geçen süre değerleri incelendiğinde idealizm, yaratıcılık ve realizm sunum videolarında, puanlayıcılar ilk olarak sunum yapan öğrenciye odaklanmıştır. Puanlayıcıların ilk sabitlendikleri ilgi alanı ile en fazla sabitlendikleri ilgi alanı birbiriyle paralellik göstermektedir.

Puanlayıcılar arasında Ortalama sabitlenme süreleri incelendiğinde sunum yapan birey ilgi alanı, sunum yapılan tahta ilgi alanına göre bütün sunumlarda daha yüksek değere

sahip olduğu bulunmuştur. Ortalama sabitleme süresi metriği puanlayıcıların dikkatinin yoğunlaştığı noktayı belirtmektedir (Jian, 2016). Puanlayıcıların dikkatlerini sunum yapan bireylere verdikleri yorumu yapılabilmektedir.

Sabitlenme sayıları da yine puanlayıcıların bilişsel çabalarının bir göstergesi olarak yorumlanabilmektedir (Rayner ve ark., 2006). Sabitleme sayısı metriği incelendiğinde, yaratıcılık sunumunda puanlayıcılar, sunum yapan öğrenci ilgi alanı için $P2 > P1 > P3$ şeklinde sıralanmaktadır. İlgi alanı üzerinde toplam sabitleme süresi için ise puanlayıcılar, $P3 > P2 > P1$ şeklinde sıralanmaktadır. Sabitleme sayısının artması puanlayıcının daha fazla çaba gösterdiği şeklinde yorumlanmaktadır.

Tartışma

Puanlayıcılar arasındaki uyum, Cohen Kappa analiziyle zayıf ve orta derecede belirlenmiştir (Landis & Koch,1977). Kappa istatistiği puanlayıcılar arası uyumu etkileyen hata kaynakları hakkında bilgi verememektedir. Puanlayıcılar arasındaki uyum, Fleiss Kappa ile incelendiğinde puanlayıcılar arasında orta derecede uyum olduğu belirlenmiştir (Landis & Koch,1977). Sınıf İçi Korelasyon Katsayısı ile elde edilen sonuçları, İdealizm ve Realizm sunumlarında, puanlayıcılar arası uyumun iyi olduğu görülmektedir Yaratıcılık sunumu için puanlayıcılar arasındaki uyumun zayıf olduğu yorumu yapılabilmektedir (Fleiss ve ark., 2003). Göz izleme uygulaması sonucunda elde edilen Sınıf İçi Korelasyon değerleri, sınıf içi uygulama değerlerine göre daha yüksektir.

Genellenebilirlik kuramına göre değişkenlik kaynakları; sunu (%39,60), U*P*S*M ortak etkisi (%27), madde (%11,20), S*M ortak etkisi (%8,50), P*M ortak etkisi (%6,10) şeklinde sıralanmaktadır. Bu sonuçlar sunuma göre puanlamanın farklılaştığı şeklinde yorumlanmaktadır. Puanlayıcı değişkenlik kaynağının %0 olması puanlayıcılar arasında farklılık bulunmadığı, puanlamaların tutarlı yapıldığı şeklinde yorumlanabilir. Çok yüzeyli Rasch analizine göre puanlayıcıların birbirleriyle uyumlu puan verdikleri yorumu yapılabilmektedir. Çok yüzeyli Rasch analizine göre, puanlama açısından uygulamalar

arasında anlamlı bir farklılık bulunmadığı yorumu yapılabilir. Genellenebilirlik sonuçlarına göre de değişkenlik kaynakları arasında uygulama ana etkisi %0 olarak bulunmuştur. Bu şekilde genellenebilirlik ve çok yüzeyli Rasch analizi birbirini desteklemektedir. Yapılan çalışmalarda bu iki model sonuçlarının birbirini desteklediği belirtilmiştir (Güler; 2008; Şata, 2019; Yıldırım Seheryeli, 2018).

Çok yüzeyli Rasch analizine göre puanlayıcılar arasında katılık ve cömertlik açısından anlamlı bir fark olduğu yorumu yapılabilmektedir. Yapılan çalışmalarla benzer sonuçların bulunduğu söylenebilir (İlhan, 2015; Tobaş, 2020; Şata, 2019).

Çok yüzeyli Rasch analizine göre puanlayıcılar arası güvenilirlik Kappa istatistiği 0,01 olarak hesaplanmıştır. Bu değer puanlayıcılar arasında zayıf bir uyum olduğunu göstermektedir (Landis & Koch,1977). Böylece klasik uyum istatistikleri ile paralel bir sonuç elde edilmiştir.

Çok yüzeyli Rasch analizine göre, sunu konusu değişikçe puanlamalar arasında anlamlı bir farklılığın olduğu görülmüştür. Bu değer genellenebilirlik sonucuyla da desteklenmektedir. Bunun yanında puanlayıcıların göz izleme metrikleri incelendiğinde, Realizm, İdealizm ve Yaratıcılık sunumlarında puanlayıcıların ortalama sabitleme süreleri, ilgi alanı üzerinde toplam sabitleme süreleri farklılık göstermektedir. Bu durum puanlayıcı sunu değişikçe puanlamalarının değişmesi durumunu destekleyen bir nitelik göstermektedir.

Çok yüzeyli Rasch analizine göre, göz izleme yöntemiyle elde edilen ortalama sabitleme sürelerine göre puanlayıcıların sunum yapan kişi ve sunumun yapıldığı tahta ilgi alanlarından en çok sunumu yapan kişiye odaklandığı görülmektedir. Ortalama sabitleme süreleri ile görevin zorluğu arasında bir ilişki olduğu yapılan çalışmalarda belirtilmiştir (Jian ,2016; Tepgeç ve Seferoğlu, 2019). Madde güçlüğü şeklinde ortalama odaklanma süresi metriği düşünülürse logit cetvelde aşağıdan yukarıya doğru arttığı şeklinde yorumlanabilmektedir (Wiseman, 2008).

Performans değerlendirme sürecinde, puanlama kriterlerinin açık ve net bir şekilde belirtildiği rubrikler ile yapılan puanlamada bile öznellik (sübjektiflik) söz konusu olabilmektedir. Bu nedenle, puanlamadaki bu özneliği azaltmada, göz izleme sonuçlarından yararlanılmaya çalışılmıştır. Göz izleme yöntem ve araçlarının performans değerlendirmede kullanılabilirliğini incelenmiştir. Bu çalışmada ele alınan problemler bütün olarak değerlendirildiğinde göz izleme tekniğiyle elde edilen göz izleme metriklerinden yararlanılarak performans değerlendirme sürecinin kalitesinin artırılabileceği, puanlayıcı davranışlarının belirlenmesinin kolaylaşacağı, varsa tutasızlıkların kaynaklarının ortaya çıkarılabileceği görülmüştür. Daha önce yapılan çalışmalar da bu yargıyı desteklemektedir (Winke & Lim, 2015; Stuart ve ark., 2018; Ma ve ark., 2022).

Çok yüzeyli Rasch analiziyle elde edilen kategori istatistiklerinde, her bir kategorinin en az on değerinde bulunması ve ortalama ölçümlerin monoton artışı ve uyum dışı değerlerin 0,5 ile 1,5 arasında değerler alması puanlama ölçeğinin uygunluğunu göstermektedir (Linacre, 2018). Çok yüzeyli Rasch analizinde elde edilen bu sonuç, çok yüzeyli Rasch analizi uygulanan ölçeklerin geçerlik ve güvenilirlikleri ile ilgili bir yorum yapma imkanı vermektedir.

Bu çalışmanın alt amaçlarından biri puanlayıcıların performans değerlendirmede nerelere dikkat ettiği konusunda bir yargıya varmaktır. Göz hareketi metrikleri incelendiğinde puanlayıcıların bakış örüntüleri hakkında sayısal ve görsel veri elde edilebildiği bulunmuştur. Elde edilen veriler sayesinde puanlayıcıların performans değerlendirme sürecinde, ölçme kriterlerine uygun olarak sunum yapan öğrenciye odaklandıkları yorumu yapılabilir.

Diğer bir alt amaç, göz izleme yöntem ve araçlarının kullanılabilirlik özellikleri hakkında bilgi edinmektir. Pilot uygulamada sabit bilgisayar ekranına adapte edilmiş sabit göz izleme cihazı kullanılmıştır. Puanlayıcılar bilgisayar laboratuvarında, değerlendirmelerini tamamlamışlardır. Asıl uygulama da ise taşınabilir göz takip cihazı ile puanlayıcılar bulunduğu ortamda laptopdan değerlendirmelerini tamamlamışlardır. Taşınabilir göz izleme cihazının kullanım açısından daha kolay ve rahat olduğu yorumu yapılabilmektedir. Göz

izleme teknolojisinin arařtırmalarda kullanımının avantajlarının yanı sıra, pahalı bir teknoloji olması arařtırmacılar için önemli bir deavantaja neden olmaktadır. Eđer herkes için ulaşımı kolay bir teknoloji olmazsa elde edilen avantajlardan yararlanmak da mümkün olmayacaktır.

Göz izleme yöntemiyle elde edilen metriklerle, farklı performans puanlarının güvenilirliğini deęerlendirmektir. Göz izleme teknięiyle elde edilen ortalama sabitleme metrięine çok yüzeyli Rasch analizi uygulanmıştır. Bunun sonucunda puanlayıcılar arasında ortalama sabitleme süresine göre anlamlı bir farklılık olmadığı yorumu yapılmıştır. Gözlenen uyum ve beklenen uyum deęerleri incelendięinde puanlayıcılar arasında düşük bir uyum olduğu yorumu yapılabilmektedir.

Arařtırmada ele alınan dięer bir alt amaç çok yüzeyli Rasch modeli ve genellenebilirlik modelinin puanlayıcı güvenilirliğini belirlemede verdikleri parametrelerle göz izleme sonuçlarının ilişkilerini belirlemektir. Genellenebilirlik analizine göre uygulama, puanlayıcı, sunu ve madde deęişkenlik kaynakları incelendięinde toplam varyansın %39,6'sının sunudan geldięi bulunmuştur. Bunun yanında puanlayıcıların sınıf içi ve göz izleme teknięiyle verdikleri puanlar çok yüzeyli Rasch analizi sonucuna göre sunu konusu deęiştikçe puanlamalar arasında anlamlı bir farklılık bulunmuştur. Sınıf içi ve göz izleme sonucunda sunudan sunuya verilen puanlar deęişirken elde edilen sonuçların deęişmedięi görülmektedir. Göz izleme ve sınıf içi uygulamada verilen puanlar benzer olmasına rağmen sunuların kendi içinde (güçlük düzeyi, içerik, sunum süresi vb.) farklılıkları olduğundan sonuçlar deęişmektedir.

Göz izleme çalışması sonucunda, her bir sunumda puanlayıcıların odaklandığı ilgi alanları çalışmada yer alan sunum deęerlendirme dereceli puanlama anahtarına göre iki alan olarak belirlenmiştir. Bunlar sunum yapan birey ve sunum yapılan tahtadır. Sunum deęerlendirme ölçeęi, sunumun içerięini deęil sunum yapan bireyin sunum yeteneęini ölçen bir özellik göstermektedir. Bu nedenle çalışmada yer alan puanlayıcıların sunum yapan bireyin ilgi alanına daha fazla odaklanması ve bu alan üzerinde daha fazla zaman geçirmesi beklenmektedir. Göz izleme sonucu elde edilen metrikler incelendięinde genel olarak

sonular bu beklentiyi karřılar niteliktedir. Bu sonular ile puanlayıcıların ölçölmek istenilen özelliklere odaklandığı sonucunda ulaşabiliriz. Her bir puanlayıcının, her bir görev için benzer metriklerinin olması da alıřmada odaklanılan puanlayıcı güvenilirliğı için önemli bir bulgu niteliğindedir.

Performans deęerlendirme sürecinde puanlayıcılar arasındaki uyumun belirlenmesi için istatistiksel yöntemler kullanılmaktadır. Bunun yanında puanlayıcıların biliřsel süreçleri hakkında toplanan veriler, puanlayıcılar arası uyum istatistiklerle birlikte deęerlendirildiğinde, puanlayıcılar arası uyum, performans deęerlendirme sürecinin kalitesini artırmak için kullanılabilir (Engelhard ve ark., 2018; Friedman ve ark., 2022). Böylece performans deęerlendirme sürecinde puanlayıcılar arasındaki uyum hakkında yeni bir bakıř açısı geliştirilmektedir.

Tüm sonular bir arada deęerlendirildiğinde puanlayıcılar arasındaki uyumun yeterince yüksek olmadığı söylenebilir. Bu da performans deęerlendirmelerde puanlayıcılar arasında tutarlığı saęlamanın güç olduğuna ilişkin literatürü destekler niteliktedir (Ballard, 2017; Engelhard, 1996; řata, 2019; Stuart ve ark., 2018).

Bölüm 5

Sonuç ve Öneriler

Bu tez çalışması kapsamında elde edilen sonuçlar araştırma sorularına göre oluşturulmuş ve alt problemlere bağlı olarak maddeler halinde özetlenmiştir.

Araştırmada betimsel istatistikler incelendiğinde hem sınıf içi hem de göz izleme uygulamalarına göre puanlayıcıların verdiği puanların ortalama değerlerinin birbirine yakın olduğu ve normale yakın bir dağılım gösterdiği belirlenmiştir.

1. Araştırmanın birinci alt problem doğrultusunda sınıf içi ve Göz izleme ile ekran üzerinden, uzmanların değerlendirme sonuçları arasında uyum Cohen Kappa ve Fleiss Kappa katsayıları ile incelenmiştir. Bunun yanında sınıf içi ve Göz izleme uygulamalarında puanlayıcıların puanları Sınıf içi korelasyon katsayısı ile incelenmiştir.

- Kappa uyumu genel olarak düşük ve orta düzeyde bulunmuştur. Göz izleme sonuçlarına ilişkin kappa sonuçları sınıf içi değerlendirmelere ilişkin kappa sonuçlarından daha yüksek görünmektedir.
- Fleiss Kappa uyumu incelendiğinde, orta derecede bir uyum olduğu bulunmuştur. Ayrıca sınıf içi uygulamaya göre, göz izleme sonuçları için elde edilen değerlerin daha yüksek olduğu görünmektedir. Cohen Kappa da olduğu gibi Fleiss Kappa uyum değerleri incelendiğinde en düşük uyumun Yaratıcılık sunumunda, en yüksek uyumun İdealizm sunumunda elde edildiği gözlemlenmiştir.
- Sınıf içi korelasyon değerleri sınıf içi uygulamaya göre göz izleme yöntemiyle ekran üzerinde elde edilen değerler için daha yüksek bulunmuştur.

2. Araştırmanın ikinci alt problemi doğrultusunda performans değerlendirme sürecinde elde edilen puanlar Genellenebilirlik Kuramıyla analiz edilmiştir.

- Genellenebilirlik sonuçlarına göre en fazla varyans açıklama oranının sunu yüzeyine ait olduğu, göz izleme ve sınıf içi değerlendirmenin varyans açıklamaya katkısı

olmadığı görülmüştür. Puanlayıcılar sınıf içi ve göz izleme teknikleriyle birbirine benzer şekilde puanlama yapmışlardır.

3. Araştırmanın üçüncü alt problemi doğrultusunda Performans değerlendirme sürecinde elde edilen puanlara çok yüzeyli Rasch modeli analizi yapılmıştır.

- Çok yüzeyli Rasch analizi ile elde edilen değişken haritası incelendiğinde puanlayıcıların katı olduğu, sınıf içi ve göz izleme değişkenlerinin benzer olduğu bulunmuştur.
- Sınıf içi ve göz izleme uygulaması puanlarının çok yüzeyli Rasch analiziyle elde edilen uygulama yüzeyi ölçüm parametreleri incelendiğinde, puanlayıcılar arasında göz izleme ve sınıf içi uygulaması sonucunda farklılık olmadığını bulunmuştur.
- Sunu konusu değiştikçe puanlamalar arasında anlamlı bir farklılığın olduğu belirlenmiştir.
- Maddeler değiştikçe puanlamalar arasında anlamlı bir farklılığın olduğu görülmüştür.
- Puanlama sonuçlarının çok yüzeyli Rasch modeline göre kategori analiz istatistikleri incelendiğinde, puanlama ölçeğinin uygun olduğunu sonucuna ulaşılmıştır.
- Puanlayıcılar arası güvenilirliğinin düşük olduğu bulunmuştur.

4. Araştırmanın dördüncü alt problemi doğrultusunda Göz izleme sonucunda puanlayıcılardan elde edilen göz izleme metriklerine çok yüzeyli Rasch modeli analizi yapılmıştır.

- Göz İzleme Uygulaması Ortalama sabitleme Süreleri dikkate alındığında; kişi ve tahta ilgi odaklarından, ortalama sabitleme süresine göre sunum yapan kişi ilgi alanı lehine, puanlayıcıların ortalama sabitleme süreleri arasında anlamlı bir farklılık olduğu sonucuna ulaşılmıştır.

- Göz İzleme uygulaması ortalama sabitlenme süreleri metriklerine göre sunu yüzeyi ölçüm parametreleri incelendiğinde sunuya göre ortalama sabitlenme süreleri arasında anlamlı bir farklılık bulunmadığı gözlemlenmiştir.
- Ortalama sabitlenme Süreleri Metrikleri sonuçlarının çok yüzeyli Rasch modeline göre kategori analiz istatistikleri sonucuna göre puanlama ölçeğinin uygun olduğunu sonucuna ulaşılmıştır.
- Puanlayıcılar arası güvenilirliğin düşük olduğu belirlenmiştir.

5. Araştırmanın beşinci alt problemi doğrultusunda Göz izleme sonucunda puanlayıcılardan göz izleme metrikleri elde edilmiştir.

- Göz İzleme Tekniğiyle, elde edilen metriklerden, ilk sabitlenmeye kadar geçen süre metriği incelendiğinde puanlayıcıların en fazla sunum yapan öğrenciye odaklandıkları sonucuna ulaşılmıştır. İlk odaklandıkları ilgi alanları ile en fazla sabitlendikleri ilgi alanı birbiriyle paralellik göstermektedir.
- Realizm sunumunda, toplam sabitlenme sürelerine göre P3 en fazla dikkati dağılan P1 ise en az dikkati dağılan puanlayıcıdır.
- Yaratıcılık sunumunda, toplam sabitlenme sürelerine göre P2 dikkati en az dağılan, P1 ise dikkati en fazla dağılan puanlayıcıdır.
- İdealizm sunumunda, toplam sabitlenme sürelerine göre P2 dikkati en fazla dağılan puanlayıcı, P1 ise dikkati en az dağılan puanlayıcıdır.
- Bu sonuçlara göre puanlayıcıların odaklanma düzeyleri arasında fark olduğu yorumu yapılabilir.

Araştırmaya Dönük Öneriler

1. Genellenebilirlik katsayısının farklı hata kaynaklarını ayırtmada etkili olduğu belirlendiğinden performans değerlendirmelerin güvenilirlik çalışmalarında genellenebilirlik tekniğinin kullanılması önerilebilir.

2. Çok yüzeyli Rasch modeli ile performans değerlendirme süresinde değerlendirme ölçütüne bağlı olarak puanlayıcılar tarafından verilen puanlar, puanlayıcı özellikleri; merkeze kayma, halo etkisi, katılık-cömertlik, tutarsızlık ve yanlılık özellikleri belirlendikten sonra puanlayıcı eğitimi verilerek bu özelliklerin ne derece değiştiği ile ilgili çalışma yapmak önerilebilir.
3. Çok yüzeyli Rasch analiziyle elde edilen kategori istatistiklerinde, her bir kategorinin en az on değerinde bulunması ve ortalama ölçümlerin monoton artışı ve uyum dışı değerlerin 0,5 ile 1,5 arasında değerler alması puanlama ölçeğinin uygunluğunu göstermektedir (Linacre, 2018). Çok yüzeyli Rasch analizinde elde edilen bu sonuç, çok yüzeyli Rasch analizi uygulanan ölçeklerin geçerlik ve güvenilirlikleri ile ilgili bir yorum yapma imkanı vermektedir. Ölçek geliştirme çalışmalarında çok yüzeyli Rasch analizini ölçeğin geçerlik ve güvenilirliğini belirlemek için kullanılabileceği önerilebilir.
4. Göz izleme yöntemiyle puanlayıcı davranışı daha büyük örneklemede; bütün puanlayıcılar için sabit sürede, toplam sabitlenme süresi, sabitlenme sayısı, ortalama sabitlenme süresi metrikleri çok yüzeyli Rasch modeli ve Genellenabilirlik analizi incelenmesi önerilebilir.
5. Göz izleme çalışması sonucunda, her bir sunumda puanlayıcıların odaklandığı ilgi alanları çalışmada yer alan sunum değerlendirme dereceli puanlama anahtarına göre iki alan olarak belirlenmiştir. Bunlar sunum yapan birey ve sunum yapılan tahtadır. Araştırmanın amacına göre daha fazla ilgi alanı belirlenerek daha detaylı bilgi toplanabileceği önerilebilir.
6. Göz izleme tekniği sadece sunu değerlendirme sürecinde değil deney düzeneği oluşturma, röportaj yapma, gezi-gözlem gibi etkinliklerin puanlanma sürecinde kullanılarak farklı durumlarda ne kadar işlevsel olduğu incelenebilir.

7. Göz izleme sonuçları; zamansal, uzamsal ve sayım odaklı ölçümler, ısı haritası, kümülatif harita ve yol haritası olarak ele alınabilmektedir. Bunların birlikte değerlendirilmesi sonuçların yorumlanmasına katkıda bulunacağından kullanılması önerilebilir.

Uygulamaya Dönük Öneriler

1. Performans değerlendirmede göz izleme yöntemi kullanılarak puanlayıcı davranışları ile puanlayıcılar arası tutarsızlıkların kaynakları ortaya çıkarılabilir.
2. Göz izleme metrikleri, incelendiğinde, puanlayıcıların davranışları hakkında bilgi verdiği belirlenmiştir. Puanlayıcı eğitiminde etkili olabilecek çalışmalar tasarlanabilir.
3. Giyilebilir göz izleme teknolojisinden yararlanılarak gerçek sınıf ortamında veya uygulama ortamında otantik değerlendirmeler yapılabilir.
4. Üstün yetenekli öğrencilerin performanslarının değerlendirilmesi tartışma konusudur. Göz İzleme Yönteminin, üstün yetenekli öğrencilerin bilişsel süreç becerilerinin daha objektif gözlemlenmesi için bir fırsat oluşturabilir.
5. Göz izleme sonuçları performans değerlendirmenin hem puanlayıcı tutarlılığı hem de puanlayıcı davranışlarını belirleme bakımından çeşitli sonuçlar ürettiğinden kullanılması yararlı olacaktır.
6. Göz izleme teknolojisinin akademik araştırmalarda kullanımının birçok avantajı olmasına rağmen, pahalı bir teknoloji olması araştırmacılar için önemli bir deavantaja neden olmaktadır. Bu nedenle daha ucuza mal edilebilecek teknolojilerin üretilmesi veya araştırmacılara mali destek verilmesi yararlı olacaktır.

Kaynaklar

- Aiken, L. R. (2000). *Psychological testing and assessment* (10. Edition). Allyn and Bacon.
- Alemdag, E.,ve Cagiltay, K. (2018). A systematic review of eye tracking research on multimedia learning. *Computers & Education*, 125, 413–428. <https://doi.org/10.1016/j.compedu.2018.06.023>
- Alkan, C. (2005). *Eğitim Teknolojisi* (8. Baskı). Anı Yayıncılık.
- Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Prentice Hall/Pearson Education.
- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43(4), 561-573.
- Anglin, G. J. (1995). *Instructional technology: Past, present, and future*. Libraries Unlimited, Inc., PO Box 6633, Englewood, CO, 80155-6633.
- Armut, M. (2021). *Yabancılara türkçe öğretiminde öğrencilerin grafikli sorulardaki okuma becerilerinin göz izleme tekniği ile araştırılması*. [Yayımlanmamış doktora tezi]. Gazi Üniversitesi, Ankara.
- Arslan Mancar, S. (2019). *Performansa Dayalı Durum Belirlemede Puanlayıcılar Arası Güvenirlik Tekniklerinin Karşılaştırılması*. [Yayımlanmamış yüksek lisans tezi]. Ankara Üniversitesi, Ankara.
- Atılğan, H. (2004). *Genellenebilirlik kuramı ve çok değişkenlik kaynaklı Rasch modelinin karşılaştırılmasına ilişkin bir araştırma*. [Yayımlanmamış doktora tezi]. Hacettepe Üniversitesi, Ankara.
- Atılğan, H. (2008). Using generalizability theory to assess the score reliability of the special ability selection examinations for music education programs in higher education. *International Journal of Research & Method in Education*, 31(1), 63-76.

- Ballard, L. (2017). *The effects of primacy on rater cognition: An eye-tracking study*. (Master thesis). Michigan State University, USA
- Barkaoui, K. (2008). *Effects of scoring method and rater experience on ESL essay rating processes and outcomes*. (Unpublished Doctoral Dissertation). University of Toronto, Canada.
- Bartko, J. J. (1966). The intraclass correlation coefficient as a measure of reliability. *Psychological reports*, 19(1), 3-11.
- Baykul, Y. (1999). *İstatistik: Metodlar ve uygulamalar*. Ankara: Anı Yayıncılık.
- Bıkmaz Bilgen, Ö., ve Doğan, N. (2017). Puanlayıcılar Arası Güvenirlik Belirleme Tekniklerinin Karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 8(1), 63-78.
- Brennan, R. L., Gao, X., & Colton, D. A. (1995). Generalizability analyses of work keys listening and writing tests. *Educational and Psychological Measurement*, 55(2), 157-176.
- Brennan, R. L. (1992). Generalizability theory. *Educational Measurement: Issues and Practice*, 11(4), 27-34.
- Büyüköztürk, Ş., Çokluk, Ö.ve Köklü N. (2013). *Sosyal bilimler İçin istatistik*. Ankara: Pegem Akademi.
- Büyüköztürk, Ş., Akgün, Ö. E., Karadeniz, Ş. Demirel, F. ve Kılıç, E. (2016). *Bilimsel Araştırma Yöntemleri*. Ankara: Pegem Akademi.
- Carroll, C. (2006). Enhancing reflective learning through role-plays: The use of an effective sales presentation evaluation form in student role-plays. *Marketing Education Review*, 16(1), 9-13.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37-46.

- Cohen, J. (1968). Multiple regression as a general data-analytic system. *Psychological Bulletin*, 70(6, Pt.1), 426–443. <https://doi.org/10.1037/h0026714>
- Cohen. J. R., Swerdlik E. M. & Phillips, S. M. (1996). *Psychological testing and assessment*. (3th ed). London: Mayfield Publishing Company.
- Congdon, P., & McQueen, J. (2000). The stability of rater severity in large-scale assessment programs. *Journal of Educational Measurement*, 37(2), 163-178.
- Coşkun, A. (2019). *Investigation of classroom management skills by using eye-tracking technology*. [Yayımlanmamış yüksek lisans tezi]. ODTÜ , Ankara.
- Coşkuner, T. (2022). *Göz izleme yönteminden elde edilen ölçümler ile test ve madde istatistikleri arasındaki ilişkilerin incelenmesi*. [Yayımlanmamış doktora tezi]. Hacettepe Üniversitesi, Ankara.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart and Winston, 6277 Sea Harbor Drive, Orlando, FL 32887.
- Cronbach, L. J. (1972). The dependability of behavioral measurements. *Theory of generalizability for scores and profiles*, 1-33.
- Çakır, H., Çebi, A., ve Özcan, S. (2013). BÖTE nedir Nasıl tanımlanır Okul müzesiyle başlayan serüvenden insan performans teknolojilerine uzanan yolculuk. *Eğitim Teknolojisi Kuram ve Uygulama*, 3(2).
- Çokluk, O., Şekercioğlu, G., ve Büyüköztürk, S. (2014). Sosyal bilimler için çok değişkenli istatistik: SPSS ve LISREL uygulamaları. Ankara: Pegem Akademi.
- Deliceoğlu, G. (2009). *Futbol yetilerine ilişkin dereceleme ölçeğinin genellenebilirlik ve klasik test kuramı dayı güvenirliklerinin karşılaştırılması*. [Yayımlanmamış doktora tezi]. Ankara Üniversitesi, Ankara.
- Dessus, P., Cosnefroy, O., & Luengo, V. (2016, September). "Keep your eyes on'em all!": A Mobile eye-tracking analysis of teachers' sensitivity to students. *In European*

- conference on technology enhanced Learning (72-84)*. Springer, Cham.on
Technology Enhanced Learning.
- Dođan, N. (2020). *Eđitimde ölçme ve deđerlendirme*. Pegem Akademi Yayıncılık
- Dođru, Ő.C. (2019). *Karma testlerin psikometrik özelliklerini belirlemede klasik test kuramı ve rasch modelinin karşılaştırılması*. [Yayımlanmamış yüksek lisans tezi]. Hacettepe Üniversitesi, Ankara.
- Duchowski, A. T. (2017). *Eye tracking methodology: Theory and practice*. Springer.
- Dunbar, N. E., Brooks, C. F., & Kubicka-Miller, T. (2006). Oral communication skills in higher education: Using a performance-based evaluation rubric to assess communication skills. *Innovative Higher Education*, 31(2), 115-128.
- Durna, Y., ve Arı, F. (2016). Polinom fonksiyonları ile göz bakış yeri tespiti geliştirilmesi ve uygulaması. *Savunma Bilimleri Dergisi*, 15(2), 25-45.
- Ebel, R. L. (1951). Estimation of the reliability of ratings. *Psychometrika*, 16(4), 407-424.
- Elhan, A. H., ve Atakurt, Y. (2005). Ölçeklerin deđerlendirilmesinde niçin Rasch analizi kullanılmalıdır?. *Ankara Üniversitesi Tıp Fakóltesi Mecmuası*, 58(1), 47-50.
- Engelhard, G. (1996). Evaluating rater accuracy in performance assessments. *Journal of Educational Measurement*, 33(1), 56-70.
- Engelhard Jr, G. (2013). *Invariant measurement: Using Rasch models in the social, behavioral, and health sciences*. Routledge.
- Engelhard Jr, G., Wang, J., & Wind, S. A. (2018). A tale of two models: Psychometric and cognitive perspectives on rater-mediated assessments using accuracy ratings. *Psychological Test and Assessment Modeling*, 60(1), 33-52.
- Erdođmuş, F. U., ve Çađıltay, K. (2009). Türkiye'de eğitim teknolojileri alanında yapılan master ve doktora tezlerinde genel eğilimler. *Akademik Bilişim*, 9, 11-13.

- Farrokhi, F., Esfandiari, R., & Dalili, M. V. (2011). Applying the many-facet Rasch model to detect centrality in self-assessment, peer-assessment and teacher assessment. *World Applied Sciences Journal*, 15(11), 76-83.
- Fleiss, J. L., Levin, B., & Paik, M. C. (2003). How to randomize. *Statistical Methods for Rates and Proportions*. 3rd ed. Hoboken, NJ John Wiley & Sons, 86-94.
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5), 378.
- Friedman, L., Prokopenko, V., Djanian, S., Katrychuk, D., & Komogortsev, O. V. (2022). Factors affecting inter-rater agreement in human classification of eye movements: a comparison of three datasets. *Behavior Research Methods*, 1-11.
- Gerçek, C., Özcan, Ö., Ocağ, N., Ferhat, Ç., Berberoğlu, S., Çakır, E., ve Doğan, N. (2016). An evaluation of the effects of Combined health warnings on Cigarette packets through eye movements. *Journal of Baltic Science Education*, 15(6), 680.
- Goodrich, H. (1997). Understanding Rubrics: The dictionary may define " rubric," but these models provide more clarity. *Educational leadership*, 54(4), 14-17.
- Gölcür, Z. (2022). Çok yüzeyli rasch modeli puanlama desenlerine göre açık uçlu maddelerin puanlayıcılar arası güvenilirliklerinin karşılaştırılması. [Yayımlanmamış yüksek lisans tezi]. Gazi üniversitesi, Ankara.
- Gözükara Bağ, H. G., Karabulut, E., ve Alpar, R. (2010). 2x2 tablolarda gözlemciler/gözlemler arası uyumun değerlendirilmesi. *Hacettepe Diş Hekimliği Fakültesi Dergisi*, 34(1), 46-52.
- Güler, N. (2008). *Klasik test genellenebilirlik ve rasch modeli üzerine bir araştırma* [Yayımlanmamış doktora tezi]. Hacettepe Üniversitesi, Ankara.
- Gümüşbaş, V. G. (2019). *Taekwondo poomsae yarışmalarında görev yapan hakemlerin görsel dikkatlerinin "göz takip sistemi" ile incelenmesi*. [Yayımlanmamış yüksek lisans tezi]. Marmara Üniversitesi, İstanbul.

- Haladyna, T. M. (1997). *Writing Test Items to Evaluate Higher Order Thinking*. Allyn & Bacon, 160 Gould Street, Needham Heights, MA 02194-2310.
- İlhan, M. (2015). *Standart ve slo taksonomisine dayalı rubrikler ile puanlanan açık uçlu matematik sorularında puanlayıcı etkilerinin çok yüzeyli rasch modeli ile incelenmesi*. [Yayımlanmamış doktora tezi]. Gaziantep Üniversitesi, Gaziantep.
- Jian, Y. C. (2016). Fourth graders' cognitive processes and learning strategies for reading illustrated biology texts: Eye movement measurements. *Reading Research Quarterly*, 51(1), 93-109.
- Jian, Y. C. (2017). Eye-movement patterns and reader characteristics of students with good and poor performance when reading scientific text with diagrams. *Reading and Writing*, 30(7), 1447-1472.
- Johnson, R. L., Penny, J. A., & Gordon, B. (2008). *Assessing performance: Designing, scoring, and validating performance tasks*. Guilford Press.
- Johnson, R. L., Penny, J., & Gordon, B. (2000). The relation between score resolution methods and interrater reliability: An empirical study of an analytic scoring rubric. *Applied Measurement in Education*, 13(2), 121-138.
- Karakaya, İ., (2020). Sınıf içi Korelasyon Katsayısı: Puanlayıcıların Uyumunun Bir Ölçümü. *Her Yönüyle Puanlayıcılar Arası Güvenirlik Rehberi* (pp.195-206), Ankara: Pegem Akademi.
- Karasar, N. (2000). *Bilimsel Araştırma Yöntemi* (10.Baskı). Ankara: Nobel Yayın Dağıtım.
- Kubiszyn, T., & Borich, G. D. (2016). *Educational testing and measurement*. John Wiley & Sons.
- Kumar, R. (2005). *Research Methodology: A Step-by-Step Guide for Beginners*. London: SAGE.
- Kutlu, Ö., Doğan, D. C. ve Karakaya, İ. (2009). *Öğrenci başarısının belirlenmesi: performans ve portfolyaya dayalı durum belirleme*. Ankara: Pegem Akademi.

- Lai, M. L., Tsai, M. J., Yang, F. Y., Hsu, C. Y., Liu, T. C., Lee, S. W. Y., ... & Tsai, C. C. (2013). A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational research review*, 10, 90-115.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.
- Langan, A. M., Wheeler, C. P., Shaw, E. M., Haines, B. J., Cullen, W. R., Boyle, J. C., Penney, D., Oldekop, J. A., Ashcroft, C., Lockey, L., & Preziosi, R. F. (2005). Peer assessment of oral presentations: Effects of student gender, university affiliation and participation in the development of assessment criteria. *Assessment and Evaluation in Higher Education*, 30(1), 21-34. <https://doi.org/10.1080/0260293042003243878>
- Langan, A., Shuker, D.M., Cullen, W., Penney, D.N., Preziosi, R.F., & Wheeler, C.P. (2008). Relationships between student characteristics and self-, peer and tutor evaluations of oral presentations. *Assessment & Evaluation in Higher Education*, 33, 179 - 190.
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel psychology*, 28(4), 563-575.
- Lee, G., & Frisbie, D. A. (1999). Estimating reliability under a generalizability theory model for test scores composed of testlets. *Applied Measurement in Education*, 12(3), 237–255. https://doi.org/10.1207/S15324818AME1203_2 .
- Linacre, J. M. (2006). Rasch analysis of rank-ordered data. *Journal of Applied Measurement*, 7(1), 129-139.
- Linacre, J. M. (2018). A user's guide to Winsteps. Chicago, IL: Winsteps.com
- Ma, W. & Winke, P. (2022). An investigation of the impact of jagged profile on L2 speaking test ratings: Evidence from rating and eye-tracking data. *Language Assessment Quarterly*. 19(4), 394-421. <https://doi.org/10.1080/15434303.2022.2078720>
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47(2), 149–174. <https://doi.org/10.1007/BF02296272>

- Mertler, C. A. (2000). Designing scoring rubrics for your classroom. *Practical assessment, research, and evaluation*, 7(1), 25.
- Morgan, C., & O'Reilly, M. (2020). *Assessing open and distance learners*. Routledge.
- Myford, C., & Wolfe, E. (2004). Detecting and Measuring Rater Effects Using Many-Facet Rasch Measurement: Part II. *Journal of Applied Measurement*, 5, 189-227.
- Nakhaeizadeh, S., Morgan, R. M., Olsson, V., Arvidsson, M., & Thompson, T. (2020). The value of eye-tracking technology in the analysis and interpretations of skeletal remains: A pilot study. *Science & justice : journal of the Forensic Science Society*, 60(1), 36–42. <https://doi.org/10.1016/j.scijus.2019.08.005>
- Nalbantođlu, F. (2009). *Genellenebilirlik kuramında dengelenmiş ve dengelenmemiş desenlerin karşılaştırılması -intramuskuler enjeksiyon yapma istasyonu verileri üzerinde bir uygulama-*. [Yayımlanmamış doktora tezi]. Ankara Üniversitesi, Ankara.
- Nalçacıer, A. (2014). *The evaluation of visual attention on hospital personnel by "eye tracking" method*. [Yayımlanmamış yüksek lisans tezi]. Fatih Üniversitesi, İstanbul.
- Öksüzođlu, M. (2022). *Üst düzey düşünme becerilerini ölçen maddelerin öğrenci puanları ve puanlayıcı güvenirliliđi açısından incelenmesi*. [Yayımlanmamış yüksek lisans tezi]. Hacettepe Üniversitesi, Ankara.
- Özmen, E. (2021). *E-Ticaret Sitelerinde Çok Kanallı Kullanılabilirlik Ve Kullanıcı Deneyimi Çalışması*. [Yayımlanmamış doktora tezi]. Atatürk Üniversitesi, Erzurum.
- Peignot, P., & Anderson, J. R. (1999). Use of experimenter-given manual and facial cues by gorillas (*Gorilla gorilla*) in an object-choice task. *Journal of Comparative Psychology*, 113(3), 253–260. <https://doi.org/10.1037/0735-7036.113.3.253>
- Pekin, Z (2015). *Otizm Sosyal Beceriler Profili Ölçeğinde Puanlayıcılar Arası Güvenirliliđin Klasik Test Kuramı Ve Genellenebilirlik Kuramına Göre Karşılaştırılması*. [Yayımlanmamış doktora tezi]. Hacettepe Üniversitesi, Ankara.

- Popham, W. J. (1997). What's Wrong--and What's Right--with Rubrics. *Educational leadership*, 55(2), 72-75.
- Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading*, 10(3), 241-255.
- Rentz, J. O. (1987). Generalizability theory: a comprehensive method for assessing and improving the dependability of marketing measures. *Journal of Marketing Research*, 24(1), 19-28.
- Romagnano, L. (2001). Implementing the assessment standards: The myth of objectivity in mathematics assessment. *The Mathematics Teacher*, 94(1), 31-37.
- Sağlam, Z. (2022). *Çevrimiçi eğitimde üstbilişsel rehberliğin bilgi işlemsel düşünme becerisi, üst bilişsel düşünme becerisi ve programlama becerisi öz yeterlilik algısına etkisi*. [Yayımlanmamış yüksek lisans tezi]. Bartın Üniversitesi, Bartın.
- Seheryeli Yıldırım, M. (2018). *Yazılı anlatım becerisi puanlama anahtarının güvenilirliğinin klasik test, genellenebilirlik ve madde tepki kuramlarına göre incelenmesi*. [Yayımlanmamış doktora tezi]. Gazi Üniversitesi, Ankara.
- Shavelson, R. J., & Webb, N. M. (1991). Generalizability Theory: A primer. Newbury Park, CA: Sage.skills. *Innovative Higher Education*, 31(2), 2006, 115-128.
- Stuart, S., Hunt, D., Nell, J., Godfrey, A., Hausdorff, J. M., Rochester, L., & Alcock, L. (2018). Do you see what I see? Mobile eye-tracker contextual analysis and inter-rater reliability. *Medical & biological engineering & computing*, 56(2), 289–296. <https://doi.org/10.1007/s11517-017-1669-z>
- Şata, M. (2019). *Performans değerlendirme sürecinde puanlayıcı eğitiminin puanlayıcı davranışları üzerindeki etkisinin incelenmesi*. [Yayımlanmamış doktora tezi]. Gazi üniversitesi, Ankara.

- Taşdelen, G. (2009). *Nedelsky ve angoff standart belirleme yöntemlerinin genellenebilirlik kuramı ile karşılaştırılmasına ilişkin bir araştırma*. [Yayımlanmamış yüksek lisans tezi]. Hacettepe Üniversitesi, Ankara.
- Tepgeç, M., ve Seferoğlu, S. S. (2019). Öğrenme-öğretme süreçlerinin değerlendirilmesinde göz izleme yönteminin kullanımıyla ilgili bir içerik analizi çalışması. *EJER Congress 2019 Bildiri Kitabı*.
- Tobaş, C. (2020). *Performansın Değerlendirilmesinde Farklılaşan Puanlayıcı Davranışlarının Çok Yüzeyle Rasch Ölçme Modeli İle İncelenmesi*. [Yayımlanmamış yüksek lisans tezi]. Gazi Üniversitesi, Ankara.
- Töngel, E., Aydın, A., Kara, M., Çakır, R. (2020). "Bilgisayar ve öğretim teknolojileri" ve "eğitim teknolojileri" alanlarında yazılan yüksek lisans ve doktora tezlerinin araştırma eğilimleri: 2013-2018 döneminin bir görüntüsü. *Ondokuz Mayıs Üniversitesi Eğitim Fakültesi Dergisi*, 39(1), 69-82. DOI: 10.7822/omuefd.552656
- Turan, C. (2018). *An eye-tracking investigation of attachment preferences to relative clauses in turkish*. [Yayımlanmamış yüksek lisans tezi]. Hacettepe Üniversitesi , Ankara.
- Turgut, M. F. ve Baykul, Y. (2014). *Eğitimde ölçme ve değerlendirme metotları*. Ankara: Pegem Akademi Yayıncılık.
- Utku, Ö. (2019). *Investigating the effects of global reading strategy training on l2 reading comprehension through eye tracking*. [Yayımlanmamış yüksek lisans tezi]. Anadolu Üniversitesi, Eskişehir.
- Veneziano L., & Hooper J. (1997). "A method for quantifying content validity of health-related questionnaires". *American Journal of Health Behavior*, 21(1):67-70.
- Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. *Fam med*, 37(5), 360-363.

- Wang, J., Engelhard, G., Raczynski, K., Song, T., & Wolfe, E. D. (2017). Evaluating rater accuracy and perception for integrated writing assessments using a mixed-methods approach. *Assessing Writing*, 33, 36–47.
- Wang, X., Lin, L., Han, M., & Spector, J. M. (2020). Impacts of cues on learning: Using eye-tracking technologies to examine the functions and designs of added cues in short instructional videos. *Computers in Human Behavior*, 107, 106279.
- Winke, P., & Lim, H. (2015). ESL essay cognitive raters' processes in applying the Jacobs et al. rating scale: An eye movement study. *Assessing Writing*, 25, 38–54.
- Wiseman, C.S. (2008). Investigating selected facets in measuring second language writing ability using holistic and analytic scoring methods. (Doctoral dissertation). Columbia University, New York.
- Wotschack, C. (2009). *Eye movements in reading strategies: How reading strategies modulate effects of distributed processing and oculomotor control* (Vol. 1). Universitätsverlag Potsdam.
- Yakut, Y. (2021). *Fizyoterapi ve rehabilitasyon öğrencilerinde eklem hareket ölçüm becerilerinin pratik değerlendirilmesinin genellenebilirlik kuramına göre incelenmesi*. [Yayımlanmamış yüksek lisans tezi]. Hasan Kalyoncu Üniversitesi, Gaziantep.
- Yelboğa, A., ve Tavşancıl, E. (2010). Klasik test ve genellenebilirlik kuramına göre güvenilirliğin bir iş performansı ölçeği üzerinde incelenmesi. *Kuram ve Uygulamada Eğitim Bilimleri*, 10(3), 1825-1854.
- Yılmaz, K. (2019). Zarar verici sosyal davranışlar ölçeğinin puanlayıcılar arası güvenilirliğinin klasik test kuramı ve genellenebilirlik kuramına göre karşılaştırılması. [Yayımlanmamış yüksek lisans tezi]. Abant İzzet Baysal Üniversitesi, Bolu.
- Yurdugül, H. (2005). Ölçek geliştirme çalışmalarında kapsam geçerliği için kapsam geçerlik indekslerinin kullanılması. *XIV. Ulusal Eğitim Bilimleri Kongresi*, 1, 771-774.

EK-A: Sunum Değerlendirme Dereceli Puanlama Anahtarı

Ölçütler	TAMAMEN YETERLİ 4	YETERLİ 3	KISMEN YETERLİ 2	YETERLİ DEĞİL 1
GİRİŞ				
KONUVA HAZIRLIKLI GELMESİ (KONUVA HÂKİMİYET)				
ANLAŞILIR BİR SUNUM HAZIRLAMASI				
ETKİLİ VE ÖZGÜN GİRİŞ YAPMASI (DİKKAT ÇEKMESİ)				
SES TONU VE BEDEN DİLİNİ DOĞRU KULLANMASI				
GRUP ARKADAŞLARIYLA VE DİNLEYİCİLERLE İŞ BİRLİĞİ YAPMASI (SÜRECE DÂHİL ETMESİ)				
YARDIMCI EĞİTSEL MATERYALLER KULLANMASI				
SUNUMUN DİNLEYİCİLERİN DÜZEYİNE UYGUN OLMASI				
SÜREÇ				
İLGİLİ KAYNAKLARDAN YARARLANMASI				
KONU KAPSAMININ YETERLİ OLMASI				
KONULAR ARASI GEÇİŞLERİN DÜZENLİ OLMASI				
GÜNCEL ÖRNEKLERE YER VERMESİ				
SORULARA CEVAP VEREBİLMESİ				
FARKLI GÖRÜŞLERİ KARŞILAŞTIRMASI				
ÖĞRETİCİ SORULAR SORMASI				
SONUÇ				
FARKLI GÖRÜŞLERİ ÇÖZÜMLEMESİ				
ÖNEMLİ NOKTALARI ÖZETLEMESİ				
KONUVA İLGİLİ SONUÇ DEĞERLENDİRMESİ YAPMASI				
KAPANIŞI YAPMASI				
ZAMANI ETKİN KULLANMASI				

EK-B: Etik Komisyonu Onay Bildirimi



T.C.
HACETTEPE ÜNİVERSİTESİ
Rektörlük

Sayı : 35853172/

433-3125

21 Eylül 2017

EĞİTİM BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

İlgi: 18.08.2017 tarih ve 1720 sayılı yazınız.

Enstitünüz Eğitim Bilimleri Anabilim Dalı Eğitimde Ölçme ve Değerlendirme Bilim Dalı doktora programı öğrencilerinden **Mine ZORLU**'nun **Prof. Dr. Nuri DOĞAN** danışmanlığında yürüttüğü "**Performans Değerlendirmede Puanlayıcılar Arası Uyum ile Göz İzleme Sonuçları Arasındaki İlişkinin İncelenmesi**" başlıklı tez çalışması, Üniversitemiz Senatosu Etik Komisyonunun **12 Eylül 2017** tarihinde yapmış olduğu toplantıda incelenmiş olup, etik açıdan uygun bulunmuştur.

Bilgilerinizi ve gereğini rica ederim.

Prof. Dr. Rahime M. NOHUTCU
Rektör a.
Rektör Yardımcısı

EK-C: Etik Beyanı

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

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

beyan ederim.

...../...../.....

(İmza)

Mine DEMİRBAŞ

EK-Ç: Yüksek Lisans/Doktora Tez Çalışması Orijinallik Raporu

HACETTEPE ÜNİVERSİTESİ

Eğitim Bilimleri Enstitüsü

Eğitim Bilimleri Ana Bilim Dalı Başkanlığına,

Tez Başlığı : Performans Değerlendirmede Puanlayıcılar Arası Uyum İle Göz İzleme Sonuçları
Arasındaki İlişkinin İncelenmesi

Yukarıda başlığı verilen tez çalışmamın tamamı (kapak sayfası, özetler, ana bölümler, kaynakça) aşağıdaki filtreler kullanılarak **Turnitin** adlı intihal programı aracılığı ile kontrol edilmiştir. Kontrol sonucunda aşağıdaki veriler elde edilmiştir:

Rapor Tarihi	Sayfa Sayısı	Karakter Sayısı	Savunma Tarihi	Benzerlik Oranı	Gönderim Numarası
02/01 /2023	88	107564	23/12 /2022	%10	1987902446

Uygulanan filtreler:

- Kaynaklar hariç
- Alıntılar dâhil
- 5 kelimedenden daha az örtüşme içeren metin kısımları hariç

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Uygulama Esasları'nı inceledim ve çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan eder, gereğini saygılarımla arz ederim.

Ad Soyadı: Mine DEMİRBAŞ

Öğrenci No.: 11262485

Ana Bilim Dalı: Eğitim Bilimleri /Eğitimde Ölçme ve Değerlendirme

Programı: Bütünleşik Doktora

Statüsü: Y.Lisans Doktora Bütünleşik Dr.

İmza

DANIŞMAN ONAYI

UYGUNDUR.

Prof.Dr. Nuri DOĞAN

EK-D: Thesis/Dissertation Originality Report

HACETTEPE UNIVERSITY

Graduate School of Educational Sciences

To The Department of Educational Sciences

Thesis Title: Investigation Of Interrater Agreement And Eye Tracking In Performance Evaluation

The whole thesis that includes the *title page, introduction, main chapters, conclusions and bibliography section* is checked by using **Turnitin** plagiarism detection software take into the consideration requested filtering options. According to the originality report obtained data are as below.

Time Submitted	Page Count	Character Count	Date of Thesis Defense	Similarity Index	Submission ID
02/01 /2023	88	107564	23/12 /2022	%10	1987902446

Filtering options applied:

1. Bibliography excluded
2. Quotes included
3. Match size up to 5 words excluded

I declare that I have carefully read Hacettepe University Graduate School of Educational Sciences Guidelines for Obtaining and Using Thesis Originality Reports; that according to the maximum similarity index values specified in the Guidelines, my thesis does not include any form of plagiarism; that in any future detection of possible infringement of the regulations I accept all legal responsibility; and that all the information I have provided is correct to the best of my knowledge.

I respectfully submit this for approval.

Name Lastname: Mine DEMİRBAŞ
Student No.: 11262485
Department: Educational Sciences
Program: Educational Measurement And Evaluation
Status: Masters Ph.D. Integrated Ph.D.

Signature

ADVISOR APPROVAL

APPROVED
Prof. Dr. Nuri DOĞAN

EK-E: Yayınlama ve Fikrî Mülkiyet Hakları Beyanı

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kâğıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin alınarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayınlanan "**Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge**" kapsamında tezim aşağıda belirtilen koşullar haricince YÖK Ulusal Tez Merkezi / H.Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- Enstitü/Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihinden itibaren 2 yıl ertelenmiştir. ⁽¹⁾
- Enstitü/Fakülte yönetim kurulunun gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren ... ay ertelenmiştir. ⁽²⁾
- Tezimle ilgili gizlilik kararı verilmiştir. ⁽³⁾

..... / /

(imza)

Mine DEMİRBAŞ

"Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge"

- (1) Madde 6. 1. Lisansüstü teze ilgili patent başvurusu yapılması veya patent alma sürecinin devam etmesi durumunda, tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu iki yıl süre ile tezin erişime açılmasının ertelenmesine karar verebilir.
- (2) Madde 6.2. Yeni teknik, materyal ve metotların kullanıldığı, henüz makaleye dönüşmemiş veya patent gibi yöntemlerle korunmamış ve internetten paylaşılması durumunda 3. şahıslara veya kurumlara haksız kazanç; imkânı oluşturabilecek bilgi ve bulguları içeren tezler hakkında tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulunun gerekçeli kararı ile altı ayı aşmamak üzere tezin erişime açılması engellenebilir.
- (3) Madde 7. 1. Ulusal çıkarları veya güvenliği ilgilendiren, emniyet, istihbarat, savunma ve güvenlik, sağlık vb. konulara ilişkin lisansüstü tezlerle ilgili gizlilik kararı, tezin yapıldığı kurum tarafından verilir. Kurum ve kuruluşlarla yapılan işbirliği protokolü çerçevesinde hazırlanan lisansüstü tezlerle ilişkin gizlilik kararı ise, ilgili kurum ve kuruluşun önerisi ile enstitü veya fakültenin uygun görüşü üzerine üniversite yönetim kurulu tarafından verilir. Gizlilik kararı verilen tezler Yükseköğretim Kuruluna bildirilir.
Madde 7.2. Gizlilik kararı verilen tezler gizlilik süresince enstitü veya fakülte tarafından gizlilik kuralları çerçevesinde muhafaza edilir, gizlilik kararının kaldırılması halinde Tez Otomasyon Sistemine yüklenir
*Tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu tarafından karar verilir.

