


<http://kefad.ahievran.edu.tr>

# Ahi Evran Üniversitesi

## Kırşehir Eğitim Fakültesi Dergisi

ISSN: 2147 - 1037

### What Do Academicians in the Field Think About Conducting Measurement and Evaluation Processes Face-to-Face and Online?

Yıldız Yıldırım  
Melek Gülşah Şahin

#### Article Information



CrossMark

DOI: 10.29299/kefad.1426527

Received: 27.01.2024

Revised: 17.04.2024

Accepted: 18.04.2024

#### Keywords:

Face-to-face Measurement and Assessment Process,

Distance Measurement and Assessment Process,

Measurement and Assessment,

Content Analysis

#### Abstract

This research, which is basic qualitative research, aims to examine the opinions of academics working in the field of measurement and assessment and teaching measurement and assessment courses about the measurement and assessment processes they carry out face to face and online. Within the scope of the research, "Opinion Form for Face-to-Face and Distance Measurement and Assessment Processes" was applied to 17 faculty members who worked in the education faculties of universities and conducted both face-to-face and distance Measurement and Assessment courses and carried out the measurement and assessment processes in these ways. The data obtained from the form was subjected to content analysis. As a result of the content analysis, four themes were found: Measurement tools, reliability, content validity, and exam security. Within the scope of face-to-face and distance measurement and assessment processes, frequencies were determined for the categories and subcategories related to these themes, and the results were interpreted. As a result of the research, it was determined that academicians preferred multiple-choice tests as measurement tools in both face-to-face and distance measurement and assessment processes. Academics have stated that the measurements have high reliability when the process is conducted face-to-face but low reliability when distantly. Although they state that content validity is high or evidence has been provided in both types of applications, the number of academics who state this for distance measurement and assessment processes is less. They reported that exam security was higher in face-to-face measurement and assessment processes. Finally, it was determined that their preferred application type was face-to-face.

### Ölçme ve Değerlendirme Süreçlerinin Yüz Yüze ve Uzaktan Yürütülmesi Hakkında Alandaki Akademisyenler Ne Düşünüyor?

#### Makale Bilgileri



CrossMark

DOI: 10.29299/kefad.1426527

Yükleme: 27.01.2024

Düzeltilme: 17.04.2024

Kabul: 18.04.2024

#### Anahtar Kelimeler:

Yüz Yüze Ölçme ve Değerlendirme Süreci,

Uzaktan Ölçme ve Değerlendirme Süreci,

Ölçme ve Değerlendirme, İçerik Analizi

#### Öz

Temel nitel araştırma niteliğindeki bu çalışmada ölçme ve değerlendirme alanında çalışan ve ölçme ve değerlendirme dersi veren akademisyenlerin yüz yüze ve uzaktan yürüttükleri ölçme ve değerlendirme süreçleri hakkındaki görüşlerinin incelenmesi amaçlanmıştır. Araştırma kapsamında üniversitelerin eğitim fakültelerinde görev yaparak "Ölçme ve Değerlendirme" dersini hem yüz yüze hem uzaktan eğitim sürecinde yürüterek ölçme ve değerlendirme süreçlerini de bu yollarla gerçekleştiren 17 öğretim üyesine "Yüz Yüze ve Uzaktan Ölçme ve Değerlendirme Süreçlerine Yönelik Görüş Formu" uygulanmıştır. Formdan elde edilen veriler içerik analizine tabi tutulmuştur. İçerik analizi sonucunda ölçme araçları, güvenilirlik, kapsam geçerliği ve sınav güvenliği olmak üzere dört tema bulunmuştur. Yüz yüze ve uzaktan ölçme ve değerlendirme süreçleri kapsamında bu temalara ilişkin kategori ve alt kategoriler için frekanslar belirlenerek sonuçlar yorumlanmıştır. Araştırmanın sonucunda akademisyenlerin hem yüz yüze hem uzaktan ölçme ve değerlendirme süreçlerinde ölçme aracı olarak çoktan seçmeli testleri tercih ettikleri belirlenmiştir. Akademisyenler sürecin yüz yüze yürütüldüğünde ölçümlerin yüksek güvenilirlikte, uzaktan yürütülmesinde ise düşük güvenilirlikte olduğunu ifade etmişlerdir. Kapsam geçerliğinin ise her iki uygulama türünde de yüksek ya da kanıt sunulmuş olduğunu belirtmelerine rağmen uzaktan ölçme ve değerlendirme süreçleri için bunu belirten akademisyen sayısı daha azdır. Sınav güvenliğinin ise yine yüz yüze ölçme ve değerlendirme süreçlerinde daha yüksek olduğunu bildirmişlerdir. Son olarak tercih ettikleri uygulama türünün yüz yüze olduğu tespit edilmiştir.

**Sorumlu Yazar :** Melek Gülşah Şahin, Prof. Dr., Gazi Üniversitesi, Türkiye, mgulsahsahin@gazi.edu.tr, ORCID ID: 0000-0001-5139-9777.

**Yazar2:** Yıldız Yıldırım, Dr. Öğr. Üyesi, Aydın Adnan Menderes Üniversitesi, Türkiye, yildizyldirm@gmail.com, ORCID ID: 0000-0001-8434-5062.

**Atıf için:** Yıldırım, Y. & Şahin, M. G. (2024). Ölçme ve değerlendirme süreçlerinin yüz yüze ve uzaktan yürütülmesi hakkında alandaki akademisyenler ne düşünüyor?. *Kırşehir Eğitim Fakültesi Dergisi*, 25(2), 1217-1258.

## Giriş

Teknolojinin ve bilgi çağının hızlı gelişimi ve eğitime entegre olmasıyla her geçen gün açık ve uzaktan eğitimin önemi ve kullanışlılığı da artmaktadır. Ülkemizde ve dünyada özellikle COVID-19 pandemisi döneminde uzaktan eğitim, acil uzaktan eğitim adı altında okul öncesi, ilköğretim, orta öğretim, lisans, lisansüstü gibi tüm eğitim düzeylerinde hızla yaygınlaşmıştır. Bu süreçte uzaktan eğitimin öğretim boyutu Türkiye'deki üniversitelerde senkron ve asenkron dersler yapılarak yürütülmüştür. Buradan da anlaşılacağı gibi bilgi teknolojisinin gelişiminin yanı sıra aynı zamanda toplum yaşamında meydana gelen değişmelerde uzaktan eğitimin yaygınlaşmasında etkili olmuştur. Uzaktan eğitim sürecinde öğretimin sunulması, öğrencilere geri bildirimlerin nasıl verileceği, öğrencilerin başarı düzeylerinin nasıl belirleneceği de önemlidir. Çünkü öğrenciler ile etkileşim düzeyinin belirli bir algı çerçevesinde yürütülebiliyor olması nedeniyle uygulamaların derinliğini yitirmesi veya sistemin uygulanabilirliğinin kısıtlandığı durumlar ortaya çıkabilmektedir (Karataş, 2003).

Uzaktan eğitimin mekândan bağımsızlığı üstünlüğüne (Sarı, 2020) karşın ölçme ve değerlendirme süreçleri de önemli sınırlılıklarından birisidir. Ölçme ve değerlendirme süreci yürütülen eğitim sistemine ilişkin geri bildirim mekanizmasıdır. Dolayısıyla yönetime, öğretmene, öğrenciye, veliye ve sistemin diğer paydaşlarına ilişkin dönüt sağlamada önemli bir rol üstlenen ölçme ve değerlendirme süreçlerinin planlanması, incelenmesi ve düzenlenmesi her zaman eğitim alanındaki önemli konularındandır. Özellikle uzaktan yürütülen ölçme ve değerlendirme süreçlerinde süreçle ilişkin geri bildirimler, öğrenciye verilen anlık dönütler, dönem sonu sınavlar, ödevler, sınav güvenliği üzerinde durulan konulardır (Acar Güvendir ve Özer Özkan, 2021).

Ölçme ve değerlendirme süreçleri yüz yüze veya uzaktan gerçekleştirilmesinden bağımsız olarak tanılayıcı, biçimlendirici ve düzey belirleyici olmak üzere üç amaçla yürütülmektedir (Özçelik, 2010). Tanılayıcı değerlendirmede öğretim etkinlikleri başlamadan önce eğitim sistemindeki girdilerin değerlendirilmesi söz konusudur. Biçimlendirmeye yönelik değerlendirmede ise öğretim süreci boyunca yapılmakta ve gelişimi izleme amaçlanmaktadır. Biçimlendirici değerlendirme yüz yüze eğitimde olduğu gibi uzaktan eğitim sürecinde de önemli bir yere sahiptir. Çünkü biçimlendirici değerlendirme sayesinde öğrenme eksikliklerinin belirlenerek giderilmesi amaçlanmaktadır. Ayrıca öğretim sürecinin de etkililiğine dönütler alınmakta ve gerekli iyileştirmeler de gerçekleştirilebilmektedir. Düzey belirleyici değerlendirme ise öğretim sürecinin sonunda yapılmakta ve özetleyici değerlendirme olarak da tanımlanmaktadır. Bu tür değerlendirmelerde ürün/sonuç değerlendirme yapılmaktadır. Uzaktan eğitimde tercih edilen en yaygın değerlendirme türleri arasında biçimlendirici ve düzey belirleyici değerlendirme görülmektedir (James, 2016; Karadağ, 2014; Usta, Kılınç ve Okur, 2022; Shraim, 2019). Biçimlendirici değerlendirme uzaktan eğitimde de özellikle önemlidir. Çünkü değerlendirmenin etkililiği için hem öğrenci hem öğretmenin süreçte aktif olması,

öğretmenin öğrencinin gelişimini düzenli izlemesi ve öğrencinin de kendi bilgi, beceri ve yeteneğinin farkına varması sağlanmalıdır (Ocak ve Karakuş, 2021).

Ölçme ve değerlendirme süreçlerinde öğrencilerin bilişsel becerileri bağlamında bilgi, beceri ve yetenekleri dikkate alınmaktadır. Özellikle erişil olarak adlandırılan bilgi ve becerilerin değerlendirilmesinde daha çok klasik değerlendirme yöntemleri kullanılırken, yeteneklerin değerlendirilmesinde ise performansa dayalı durum belirleme yöntemleri tercih edilmektedir. Performansa dayalı durum belirlemede kullanılan ölçme ve değerlendirme araçları ise performans görevleri, proje ve portfolyo, analitik ve bütünsel dereceli puanlama anahtarı, dereceleme ölçekleri ile kontrol listeleridir (Gültekin, 2017; Kutlu, Doğan ve Karakaya, 2010; Popham, 1999; Şahin, 2019). Ölçme ve değerlendirme araçlarının kullanılmasında klasik yöntemler ve performansa dayalı değerlendirmelerden farklı sınıflamalarda mevcuttur (Başol, 2018; Bıkmaz Bilgen, 2019; Doğan, 2009; Doğan, 2019; Kilmen, 2017; Özçelik, 2010;). Bunlardan ilki geleneksel yöntemler ve tamamlayıcı yöntemler olarak sunulan sınıflamadır. Tamamlayıcı ölçme ve değerlendirme araçları geleneksel araçlardan farklı olup, geleneksel araçların sahip olduğu sınırlılıkların tamamlayıcısı olarak değerlendirme süreçlerine dahil edilmektedir. Geleneksel ölçme ve değerlendirme araçları çoktan seçmeli, açık uçlu, kısa yanıt, eşleştirmeli ve doğru-yanlış tipi maddeler, sözlü sınavlar, yazılı sınavlar olarak sunulabilir. Ölçme araçlarının sınıflandırılmasında kullanılan bir diğer sınıflamada puanlayıcıdan kaynaklı hatanın mevcut durumuna yönelik olarak objektif ve sübjektif ölçme araçlarıdır. Çoktan seçmeli, doğru-yanlış, eşleştirmeli testler objektif testler iken sözlü ve yazılı sınavlar sübjektif testler bağlamında değerlendirilebilir. Literatürde var olan sınıflandırmalar ölçme araçlarının ölçtüğü özellikler, hazırlama, uygulama ve değerlendirme süreçleri, değerlendirmede kullanıldığı aşamalar vb. özellikler dikkate alınarak sınıflandırılmıştır. Ölçme araçlarının sınıflandırılmalarından bağımsız olarak göz önüne alınması gereken en önemli özellikler güvenilirlik, geçerlik ve kullanılabilirlik. Ölçme aracının gerçekte ölçmeyi amaçladığı özelliği ölçebilme düzeyine yönelik dikkate alınması gereken en önemli özelliği geçerliktir. Ancak ölçme araçlarından elde edilen puanların ve verilen kararların hatasızlığı ile ilgili olan güvenilirlik de araştırılması gereken bir diğer özelliktir. Ölçme aracının geçerlik ve güvenilirliğinin sağlanmasından sonra dikkate alınması gereken diğer bir özellik ise kullanılabilirlik. Ölçme araçlarının kullanılabilirliği; ekonomik olması, hazırlama-uygulama süresi, uygulama – puanlama - yorumlama kolaylığı ve hazırlayıcı - uygulayıcı - cevaplayıcı nitelikleri (Turgut, 1992) göz önüne alınarak değerlendirilmelidir.

Ölçme araçları değerlendiricilere birçok anlamda zengin bilgi sağlamakla beraber birbirlerinden ölçtüğü özellik, güvenilirlik, geçerlik ve kullanılabilirlik açısından farklılık göstermektedir. Tüm bu ölçütler göz önüne alındığında en iyi ölçme aracını belirlemek çok zordur. Bu nedenle ölçme ve değerlendirme süreçlerinin yüz yüze veya uzaktan yürütülmesinden bağımsız olarak süreçlerin güvenilirlik, geçerlik ve kullanılabilirliği belirli düzeyde sağlamaya çalışılmalıdır. Çünkü yüz yüze ve

uzaktan gerçekleştirilen ölçme ve değerlendirme süreçlerinin incelenmesi ve geliştirmeye yönelik öneriler değerlendirme sonucunda verilecek kararlar bağlamında da önemlidir.

Literatürde ölçme ve değerlendirme süreçlerinin sınıf içinde ve çevrim içi/çevrim dışı olmak üzere uzaktan yürütülmesinde değerlendirmelere yönelik çeşitli araştırmalar söz konudur. Acar Güvendir ve Özer Özkan (2021) gerçekleştirdikleri çalışmada öğretmen adaylarının görüşlerine dayalı olarak çevrim içi sınavlarda değerlendirme süreçlerinin sınıf içi sınavlarla karşılaştırılarak incelemiştir. Baran (2020) çalışmasında açık ve uzaktan eğitimde kullanılan ölçme araçlarını tanıtmış ve farklı üniversitelerdeki ölçme ve değerlendirme yaklaşımlarını incelemiştir. Bozkurt ve Uçar (2018) çevrim içi ölçme ve değerlendirme süreçlerinde güvenliği sağlamaya yönelik lisans öğrencilerinin görüşlerini alarak gerçekleştirdikleri güvenliğe yönelik gereksinimleri belirlemiştir. Karadağ (2021) gerçekleştirdiği çalışmasında çevrim içi sınavlara hazırlık sürecini tartışmış ve öneriler getirmiştir. Ocak ve Karakuş (2022) üniversite öğrencilerinin çevrim içi sınavlara yönelik tutumlarını teknik unsur, önlem alma, bireysel özellik ve sınav yapısı faktörlerine dayalı olarak belirleyen bir tutum ölçeği geliştirmişlerdir. Sarı (2020) çalışmasında uzaktan eğitimde ölçme ve değerlendirme sürecini tartışmıştır. Shraim (2019), üniversite öğrencilerinin çevrimiçi sınavlara ilişkin görüşlerini incelemiştir. Taşkıran (2021) uzaktan öğretimde yönetim, öğrenme, teknoloji ve değerlendirme boyutlarını ele alarak tartışmıştır. Usta, Kılınç ve Okur (2022) üniversitede derse giren ölçme ve değerlendirme veya açık ve uzaktan öğrenme alanında çalışan uzmanlarla odak grup görüşmeler gerçekleştirerek öğrenme ortamının açık ve uzaktan öğrenmede ölçme ve değerlendirme süreçlerinde dikkat edilmesi gereken unsurlara ilişkin görüşleri incelemiştir. Bu araştırmada literatürde var olan araştırmalardan farklı olarak ölçme ve değerlendirme alanında çalışmakta olup ölçme ve değerlendirme süreçlerini hem yüz yüze hem uzaktan yürüten akademisyenlerin bu süreçlere ilişkin görüşlerinin ortaya konulması amaçlanmıştır. Bu doğrultuda ölçme ve değerlendirme alanında çalışan akademisyenlerin her iki uygulama türüne yönelik görüşleri alınmıştır. Ölçme ve değerlendirme alanında uzman olmalarından dolayı araştırmada ele alınan problemleri uzmanlık alanları doğrultusunda derinlemesine inceleyebilecekleri düşünülmüştür. Araştırmada ayrıca araştırma sonuçlarından yola çıkılarak öğrencilerin başarılarının izlenmesi ve değerlendirilmesinde kullanılan ölçme araçlarının güvenilirlik, geçerlik, kullanılabilirlik ve sınav güvenliğini iyileştirmeye yönelik uygulayıcılara bilgi sağlayacağından önemlidir. Araştırmanın amacı doğrultusunda araştırmanın alt problemleri aşağıda sunulmuştur.

1. Akademisyenler yüz yüze ve uzaktan ölçme ve değerlendirme süreçlerinde hangi ölçme araçlarını tercih etmektedirler?

2. Akademisyenlerin yüz yüze ve uzaktan ölçme ve değerlendirme süreçlerinin güvenilirliği hakkındaki görüşleri nelerdir?

3. Akademisyenlerin yüz yüze ve uzaktan ölçme ve değerlendirme süreçlerinin kapsam geçerliği hakkındaki görüşleri nelerdir?

4. Akademisyenlerin yüz yüze ve uzaktan ölçme ve değerlendirme süreçlerinin sınav güvenliği hakkındaki görüşleri nelerdir?

5. Akademisyenlerin ölçme ve değerlendirme süreçlerinde uygulama türüne ilişkin tercihleri nelerdir?

## Yöntem

### Araştırmanın Türü

Bu araştırmada hem yüz yüze hem de uzaktan eğitimi yürüten ölçme değerlendirme alanında akademisyenlerin görüşlerine göre her iki uygulamanın araştırma kapsamında belirlenen özellikler bağlamında incelenmesi amaçlanmıştır. Araştırma var olan bir durumu/olguya ilişkin zengin tanımlamalar getirmeyi amaçladığından temel nitel bir araştırma olarak nitelendirilebilir (Meriam, 2009).

### Çalışma Grubu

Araştırma kapsamında üniversitelerin eğitim fakültelerinin ölçme ve değerlendirme anabilim dalında görev yaparak "Ölçme ve Değerlendirme" dersine yönelik ölçme ve değerlendirme süreçlerini hem yüz yüze hem uzaktan yürüten 17 akademisyene ulaşılmış ve A1, A2, A3, ..... A17 şeklinde kodlanmıştır. Akademisyenlere ilişkin demografik bilgiler Tablo 1'de sunulmuştur.

Tablo 1. Katılımcıların demografik özellikleri

Değişkenler		Frekans
Cinsiyet	Kadın	11
	Erkek	6
Ünvan	Prof. Dr.	2
	Doç. Dr.	8
	Dr. Öğr. Üyesi	5
	Arş. Gör. Dr.	2
	2-5	6
Derse girme yılı	6-10	5
	11-15	4
	16 ve üzeri	2
Yüz yüze ders verilen dönem	1-5	6
	6-10	6
	11-15	3
Uzaktan eğitimde ders verilen dönem	16 ve üzeri	2
	1-5	15
	6-10	2

Tablo 1 incelendiğinde çalışma grubunda yer alan akademisyenlerin çoğunluğu kadınlardan (f=11) oluşmaktadır. Katılımcıların ünvanları ve derse girme yılları incelendiğinde farklı özelliklere sahip oldukları belirtilebilir. Katılımcıların çoğunluğu doçent ünvanına sahip (f=8) ve 2-5 yıl arası (f=6) derse girme deneyimine sahiptir. Çalışma grubunda yer alan katılımcıların yüz yüze ve uzaktan

eğitimde ders verdikleri dönemler incelendiğinde ise yüz yüze eğitimde çoğunluk 1-5 ve 6-10 dönem (f=5), uzaktan eğitimde ise çoğunluk 1-5 dönem (f=15) derse girmiştir.

### **Veri Toplama Araçları**

Araştırma kapsamında araştırmacılar tarafından “Yüz Yüze ve Uzaktan Ölçme ve Değerlendirme Süreçlerine Yönelik Görüş Formu” geliştirilmiştir. Formun geliştirilmesi sürecinde öncelikle ölçme ve değerlendirme dersini yürüten üç araştırmacı ile görüşme yapılarak ve literatür taranarak taslak bir form oluşturulmuştur. Oluşturulan taslak forma ilişkin altı ölçme ve değerlendirme uzmanından görüş alınarak forma nihai şekli verilmiştir. Bu aşamada forma ilişkin görüşlerin her biri değerlendirilerek form düzenlenmiş ve formun her bir yeni hali başka bir uzman görüşüne sunulmuş ve bu sırayla süreç takip edilmiştir. Formun nihai hali iki bölümden oluşmaktadır. Birinci bölümde cinsiyet, ünvan, kıdem, yüz yüze ve uzaktan ölçme ve değerlendirme dersinin verildiği dönem sayısı gibi betimleyici değişkenler yer almaktadır. İkinci bölümde ise beş açık uçlu soru yer almaktadır. Beş açık uçlu soru araştırmanın problemlerini yansıtmaktadır ve ölçme ve değerlendirme alanında çalışan akademisyenlerin yüz yüze ve uzaktan ölçme ve değerlendirme süreçlerinde tercih ettikleri ölçme araçlarına, güvenilirliğe, geçerliğe, sınav güvenliğine ve tercih edilen uygulama türüne yönelik sorulardır.

### **Verilerin Analizi**

Araştırma verileri “Yüz Yüze ve Uzaktan Ölçme ve Değerlendirme Süreçlerine Yönelik Görüş Formu” ile Google Forms aracılığıyla toplanmıştır. Veri toplama aracında yer alan maddelerin yanıtlanması zorunlu olarak belirlenmiştir. Bu form ile ölçme ve değerlendirme alanındaki akademisyenlerden elde edilen veriler içerik analizine tabi tutulmuştur. Araştırmada içerik analizi her iki araştırmacı tarafından yapılmıştır. İçerik analizi sonucunda kod, alt kategori, kategori ve temalar oluşturulmuştur. (Yıldırım ve Şimşek, 2011). İçerik analizinde oluşan temalar formdaki soruların dayandığı temel problemler ve araştırmanın alt problemleriyle paralel olarak oluşmuştur. Bu nedenle bulgular bu temalar bağlamında sunulmuştur.

### **İnanırcılık, Aktarılabirlik ve Tutarlılık**

Araştırma kapsamında belirlenen kategori ve alt kategorilere ilişkin doğrudan alıntılara raporlamada yer verilmiştir. Bu inanırcılığa kanıt olarak sunulabilir. Diğer yandan çalışma grubuna dahil edilen akademisyenler ölçme ve değerlendirme alanından amaçlı olarak seçilmiştir. Çünkü ölçme ve değerlendirme süreçleri, ölçme araçları, geçerlik, güvenilirlik, sınav güvenliği vb. gibi bu araştırmada ele alınan kavramlar ölçme ve değerlendirme alanının çalışma konularındandır. Ek olarak araştırmanın çalışma grubunun oluşturulmasında farklı derse girme deneyimine sahip farklı akademik ünvanlardan akademisyenler seçilmiştir. Bu faktörler ise aktarılabirliğe kanıt olarak sunulabilir (Arslan, 2022; Yıldırım ve Şimşek, 2011).

Araştırmada içerik analizi kapsamındaki kodlamalar araştırmacılar tarafından ayrı ayrı yapılmıştır. Daha sonra bu kodlayıcıların arasındaki uyum yüzdesi Miles ve Huberman (1994)'ün uyum yüzdesine göre hesaplanmıştır. Kodlayıcılar arası uyum yüzdesi %91,18 bulunmuştur. Uyumsuzluğa neden olan kodlamalar araştırmacılar tarafından birlikte değerlendirilmiş olup çatışmanın olduğu birimlerde uzlaşma sağlanmış ve böylelikle çatışmalar çözülmüştür. Bu doğrultuda araştırma kapsamında tutarlılığa kanıt sağlandığı söylenebilir.

### Araştırmanın Etik İzinleri

Yapılan bu çalışmada "Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesi" kapsamında uyulması belirtilen tüm kurallara uyulmuştur. Yönergenin ikinci bölümü olan "Bilimsel Araştırma ve Yayın Etiğine Aykırı Eylemler" başlığı altında belirtilen eylemlerden hiçbirini gerçekleştirilmemiştir.

**Etik kurul izin bilgileri:** Etik değerlendirmeyi yapan kurul adı = Gazi Üniversitesi Etik Komisyonu

Etik değerlendirme kararının tarihi= 19.12.2023

Etik değerlendirme belgesi sayı numarası= E.830002

### Bulgular

Araştırmada yapılan içerik analizi sonucunda araştırma problemleriyle paralel olarak beş tema oluşmuştur. Bu temalar; "ölçme araçları", "güvenirlik", "kapsam geçerliği", "sınav güvenliği" ve "uygulama türü" şeklindedir. Araştırmanın bulguları bu temalar ışığında sunulmuştur.

#### i. Ölçme Araçları

Akademisyenlerin yüz yüze ve uzaktan sınavlarda tercih ettikleri ölçme araçlarına ilişkin frekanslar Tablo 2'de sunulmuştur.

Tablo 2. Akademisyenlerin tercih ettikleri ölçme araçları

Tercih edilen ölçme aracı	Yüz yüze	Uzaktan
	Frekans	Frekans
Çoktan seçmeli test	14	14
Kısa cevaplı test	1	0
Yazılı yoklama	1	2
Karma test	5	2
Ödev	3	4

Tablo 2 incelendiğinde akademisyenlerin yüz yüze sınavlarda çoğunlukla çoktan seçmeli testleri (f =14) tercih ettikleri görülmüştür. Örneğin A2 "Genellikle vize ve finaller için çoktan seçmeli test kullanmayı tercih ediyorum." şeklinde, A9 ise "Çoktan seçmeli ara sınav ve final yapıyorum." şeklinde çoktan seçmeli testleri tercih ettiklerini belirtmişlerdir. Çoktan seçmeli testlerden sonra en çok karma testlerin (f=5) tercih edildiği görülmüştür. Örneğin A15 "Farklı madde türleri içeren araçlarla öğrenme düzeylerinin hangi seviyelerde olduğunu belirlemek istiyorum." şeklinde ifade ederek karma

testleri tercih ettiğini belirtmiştir. En az tercih edilen ölçme araçları ise kısa cevaplı test (f=1) ve yazılı yoklamadır (f=1).

Akademisyenlerin uzaktan sınavlarda tercih ettikleri ölçme araçları incelendiğinde ise uzaktan sınavlarda da yüz yüze sınavlarda olduğu gibi en çok çoktan seçmeli testlerin (f=14) tercih edildiği görülmüştür. Örneğin hem çoktan seçmeli test hem yazılı yoklama kullandığını belirten A4:

Benzer durum, uzaktan eğitim içinde geçerlidir. Genellikle vize ve finaller için çoktan seçmeli test kullanmayı tercih ediyorum. Mazeret ve bütünleme sınavlarında ise klasik sınav kullanmayı tercih ediyorum. Çoktan seçmeli test tercih etmemin sebebi kullanışlı olması, çok sayıda soruyu kolayca puanlayabilmek. Bütünleme ve Mazeret sınavı için tekrar test hazırlamak zor olduğu için ise klasik sınavı tercih ediyorum.

diyerek yüz yüze ve uzaktan eğitimde benzer madde türlerinden yararlandığını belirtmiştir. Benzer olarak A10'da "Tüm sınavlarımda çoktan seçmeli maddelerden oluşan testler kullandım." şeklinde uzaktan sınavlarda da çoktan seçmeli testler uyguladığını ifade etmiştir. Uzaktan sınavlarda çoktan seçmeli testlerden en çok tercih edilen ölçme aracının ödev olduğu görülmüştür ve bu ölçme aracı uzaktan sınavda yüz yüze sınava göre daha çok tercih edilmektedir (f=4). Hem çoktan seçmeli test hem ödev kullanan A17 ölçme aracının uzaktan ya da yüz yüze uygulanmasına göre değişmediğini şu şekilde ifade etmiştir:

Uzaktan eğitimle de benzer şekilde vize ve final sınavlarında çoktan seçmeli maddelerden oluşan bir sınav yaptım ayrıca ödevde de kazanıma uygun olarak soru yazmalarını istedim. Ölçme değerlendirme araçları noktasında uzaktan ya da yüz yüze olması benim tercih ettiğim ölçme değerlendirme aracını farklılaştırmadı.

Karma testlere ilişkin bulgular incelendiğinde ise yüz yüze eğitime göre daha az tercih edildiği belirlenmiştir (f=2). Karma test uygulayan A12 "Uzaktan olan derslerde, ders başında quizler yaparak öğrencilerin konuları düzenli çalışmasını sağlıyorum. Uzaktan derslere öğrenciler aktif katılmadığı için bu yöntem işe yarıyor. Bu quizlerde farklı madde türlerine yer veriyorum" diyerek gerçekleştirdiği kısa sınavlarda madde türlerini çeşitlendirdiğini ifade etmiştir. Ek olarak yazılı yoklamanın uzaktan sınavlarda yüz yüze sınavlara göre daha çok tercih edildiği görülmekle beraber (f=2) uzaktan sınavlarda kısa cevaplı testlerin akademisyenler tarafından tercih edilmediği görülmüştür.

Akademisyenlerin yüz yüze ve uzaktan sınavlarda bu ölçme araçlarını tercih etme nedenlerine ilişkin kategori, alt kategori ve frekanslar ise Tablo 3'te verilmiştir.



Tablo 3. Akademisyenlerin tercih ettikleri ölçme araçlarını tercih etme nedenlerine kategorisine ve alt kategorilerine ilişkin frekanslar

Tercih etme nedenleri	Yüz yüze	Uzaktan
Çoktan seçmeli test	Frekans	Frekans
Kullanışlılık	8	7
KPSS’de de bu test türünün kullanılması	1	1
Kapsam geçerliği ve güvenilirlik	3	3
Objektif puanlama	2	2
Öğrenmeleri yoklama ve geri dönüt verme	2	0
Sınav Güvenliği	0	1
Yazılı yoklama		
Kullanışlılık	1	1
Sınav Güvenliği	0	1
Karma test		
Konu alanı ve bilişsel düzey çeşitliliği	1	0
Bireysel farklılıklar	1	1
Öğrenci motivasyonunu sağlamak	0	1
Sınav güvenliği	0	1
Madde türlerinin farklı üstünlükleri olması	1	0
Ödev		
Kullanışlılık	1	0
Öğrenci motivasyonunu sağlamak	0	1
Sınav güvenliği	0	1

Tablo 3 incelendiğinde akademisyenlerin yüz yüze sınavlarda en çok kullanışlılık (f=8) nedeniyle çoktan seçmeli testleri tercih ettikleri görülmüştür. Bunu A2 “Ayrıca küçük bir üniversitede çalıştığım ve genellikle ders yüküm 30 saat ve üstü olduğundan dolayı da çoktan seçmeli testleri tercih ediyorum.” şeklinde, A4 “Çoktan seçmeli test tercih etmemin sebebi kullanışlı olması, çok sayıda soruyu kolayca puanlayabilmek.” şeklinde, A10 ise “Öğrenci sayısı hep çok fazlaydı (yaklaşık 300-350 kişi)” şeklinde ifade etmiştir. Çoktan seçmeli testi tercih etme nedenlerine değinmeyen üç akademisyen bulunurken, çoktan seçmeli test kullanmasının sebebinin kapsam geçerliği ve güvenilirlikle ilişkilendiren de üç akademisyen bulunmaktadır. Kapsam geçerliği ve güvenilirlik için tercih ettiğini belirten akademisyenlerden A6 bunu “çoktan seçmeli test kullanma sebepim güvenirligin ve geçerliğin yüksek olması ve aynı zamanda objektif puanlanıyor olmasıdır.” şeklinde ifade ederken, A7 ise “ölçülmesi gereken kazanımları önemli ölçüde kapsamak ve objektif değerlendirme yapmak.” şeklinde ifade etmiştir. A6 ve A7 aynı zamanda çoktan seçmeli testlerde objektif puanlama yapılabilmesine de vurgu yapmıştır. Yüz yüze sınavlarda yazılı yoklamaları kullanan tek akademisyenin (A4) kullanışlılık nedeniyle bu ölçme aracını tercih ettiği gözlemlenmiştir. Diğer yandan karma testleri tercih etme nedenini belirten üç akademisyenin birbirinden farklı nedenler sunduğu görülmüştür. Bu nedenler, karma testlerin her konu alanı ve bilişsel düzeyi kapsamayı sağlaması, bireysel farklılıkları gözetiyor olması ve her madde türünün farklı üstünlüklerinin ve sınırlılıklarının bulunması şeklinde sıralanabilir.

Uzaktan sınavlarda ölçme araçlarının tercih edilme nedenleri incelendiğinde yüz yüze sınava benzer olarak çoktan seçmeli testlerin en çok kullanışlılık (f=7) nedeniyle tercih edildiği görülmüştür. Örneğin A1: “kalabalık gruplarda okuma kolaylığı” ve A11: “Çok fazla sayıda (en az 250 ve üzeri)

öğrenciye ders verdiğim için” diyerek sınavın uygulanacağı grupların kalabalık olmasına değinmişlerdir. Yine yüz yüze sınavlara benzer olarak çoktan seçmeli testlerin tercih edilme nedenleri arasında kullanışlılık kategorisini kapsam geçerliği ve güvenilirlik kategorisinin takip ettiği belirlenmiştir. Yüz yüze ve uzaktan sınavlarda çoktan seçmeli testleri kapsam geçerliği ve güvenilirlik nedeniyle tercih ettiğini belirten akademisyenler aynı katılımcılardır (A5, A6 ve A7). Burada göze çarpan bir bulgu sınav güvenliği nedeniyle uzaktan sınavlarda çoktan seçmeli testleri tercih eden bir akademisyenin (A1) bulunmasıdır ve şu şekilde ifade etmiştir: “Soruların ve seçeneklerin yerini değiştirebilme imkanının güvenlik endişelerini azaltması”. Yazılı yoklamayı tercih eden 2 akademisyen bulunmasıyla birlikte bu akademisyenlerden A4 yazılı sınavın hazırlanmasındaki kullanışlılığı nedeniyle mazeret ve bütünleme sınavlarında kullandığını belirtirken, A5 sınırlı sürede kopya çekilmesini en aza indirmek için bu ölçme aracını tercih ettiğini belirterek sınav güvenliğini vurgulamıştır. Diğer yandan karma testleri tercih eden iki akademisyen bulunmakla birlikte, bu akademisyenler birbirinden farklı üç neden sunmuştur. Bu nedenler, öğrenci motivasyonunu sağlamak, bireysel farklılıkları gözetiyor olması ve her madde türünün farklı üstünlüklerinin ve sınırlılıklarının bulunması şeklinde sıralanabilir. Son olarak uzaktan sınavlarda çoktan seçmeliden sonra en çok tercih edilen ölçme araçlarından ödevi tercih etme nedenini belirten iki akademisyenin (A7 ve A14) bulunmaktadır. Bu akademisyenlerin ödevi tercih etme nedenlerinin, karma testlerin tercih edilmesi nedenleriyle aynı olduğu ve öğrenci motivasyonu sağlama ile sınav güvenliği sebepleriyle kullanıldığı belirlenmiştir.

## ii. Güvenirlik

Güvenirliğe ilişkin veriler incelendiğinde akademisyenlerin güvenilirliğin mevcut durumlarına ilişkin frekanslar Tablo 4’te sunulmuştur.

Tablo 4. *Güvenirliğin mevcut durumu kategorisine ve alt kategorilerine ilişkin frekanslar*

Güvenirliğin mevcut durumu	Yüz yüze	Uzaktan
	Frekans	Frekans
Yüksek güvenilirlik	13	3
Düşük güvenilirlik	2	13

Tablo 4 incelendiğinde akademisyenlerin büyük bir çoğunluğu (f=13) yüz yüze sınavların yüksek güvenilirlik ile ölçüm yaptığını ifade etmiştir. Örneğin güvenilirliğin yüksek olduğunu belirten A2: “güvenirlik analizleri yaptığımda genellikle yüksek”, A5: “yüz yüze yapılan testlerden elde edilen ölçümlerin güvenilirliği yüksektir.” ve A12: “Oldukça yüksek” şeklinde ifade etmiştir. Ancak bunların aksine iki akademisyen güvenilirliğin düşük olduğunu bildirmiştir. Örneğin A7: “güvenirlik daha düşük ama genel olarak kötü denilmez” şeklinde ifade ederek yüz yüze sınavların uzaktan sınavlara göre daha düşük güvenilirlikle ölçümler elde ettiğini belirtmiştir. Uzaktan sınavlara ilişkin bulgular incelendiğinde akademisyenlerin büyük bir çoğunluğu (f=13) uzaktan sınavların düşük güvenilirlik ile ölçüm yaptığını ifade etmiştir. Örneğin uzaktan sınavlarda güvenilirliğin düşük olduğunu belirten A3: “Ciddi güvenilirlik ve güvenlik sorunları var”, A6: “uzaktan yapılan sınavların güvenilir olmadığını düşünüyorum.” ve

A10:“ güvenilirlik tehdit altında” şeklinde ifadeleri kullanmışlardır. Diğer yandan güvenilirliğin uzaktan sınavlarda yüksek güvenilirlikle ölçümler elde ettiğini ifade eden iki akademisyen bulunmaktadır. Örneğin karma testler kullanmayı tercih eden A15: “Puanlama güvenilirliğinin yüksek yanlılığın az olduğunu düşünüyorum.” şeklinde görüş belirtmiştir.

Güvenirliğin mevcut durumunun nedenlerine ilişkin alt kategoriler yüksek ve düşük güvenilirlik kategorileri için ayrı ayrı belirlenmiş olup alt kategorilere ait frekanslar Tablo 5’te sunulmuştur.

Tablo 5. *Güvenirliğin mevcut durumunun nedenleri kategorisine ve alt kategorilerine ilişkin frekanslar*

Güvenirliğin mevcut durumunun nedenleri	Yüz yüze	Uzaktan
	Frekans	Frekans
<b>Yüksek güvenilirlik</b>		
Sınav güvenliği	1	0
Hata kaynaklarının indirgenmesi	1	0
Soru sayısı	1	0
Madde istatistikleri bilinen soru kullanımı	1	0
Grubun heterojenliği	1	0
<b>Düşük güvenilirlik</b>		
Sınav güvenliği	2	9
Soru sayısı	0	1
Sınıfların kalabalık olması	2	0
Sahip olunan teknoloji	0	1
Şans başarısı	1	0
Grubun homojenliği	0	1

Tablo 5 incelendiğinde yüz yüze sınavlarda güvenilirliğin yüksek olmasının sebeplerine görüşlerinde yer veren beş akademisyen bulunmaktadır ve bu akademisyenlerin her birini birbirinden farklı nedenlere değindiği görülmüştür. Bu nedenler sınav güvenliği, hata kaynaklarının indirgenmesi, soru sayısı, madde istatistikleri bilinen soru kullanımı ve grubun heterojenliği alt kategorilerini oluşturmuştur. Akademisyenlerden biri (A5) sınav güvenliği yüksek olduğu için güvenilirliğin yüksek olduğunu “Kopya çekilmesinin önüne geçilebildiği için” şeklinde belirtirken, başka bir akademisyen (A6) ise hata kaynaklarını bilmesi nedeniyle tesadüfi hataları en aza indirdiğini ve bu nedenle yüz yüze sınavlarının güvenilir olduğunu “Nitekim hata kaynaklarını bilmem sebebiyle olası tesadüfi hataları en aza indirdiğimi düşünüyorum.” şeklinde belirtmiştir. Uzaktan sınavlarda ölçümlerin güvenilirliğinin yüksek olduğunu belirten akademisyenlerden herhangi birinin sebep belirtmediği görülmüştür. Yüz yüze sınavlarda güvenilirliğin düşmesinin sebeplerinin ise sınav güvenliği, sınıfların kalabalık olması, şans başarısı ve grupların homojenliği alt kategorilerinde toplandığı görülmüştür. Bu alt kategorilerden sınav güvenliği (f=2) ve sınıfların kalabalık olması (f=2) alt kategorilerinin ön plana çıktığı belirlenmiştir. Örneğin A7:“kalabalık sınıflarda kopya vs. sorunları olma ihtimalinden dolayı” diyerek her iki alt kategoride de vurgu yapmıştır. Uzaktan sınavlarda düşük güvenilirliğin nedenlerinin sınav güvenliği, soru sayısı, sahip olunan teknoloji ve grubun homojenliği şeklinde alt kategorilere ayrıldığı görülmüştür. Bu alt kategorilerdeki dağılım incelendiğinde akademisyenlerin en çok (f=9) sınav güvenliğini neden olarak gösterdiği belirlenmiştir. Örneğin A1: “Güvenlik endişeleri var. Kendisi mi

cevaplardı, yardım aldı mı, sınav sırasında internetten yararlandı mı vs.”, A10: “Sınav güvenliği/kopya gibi sebeplerden dolayı” ve “Öğrenci nerede ve nasıl sınava giriyor kontrol edilemiyor. Yanında biri olabilir. Öğrencileri birlikte olabilir. Basılı kaynaktan yararlanabilir. Çok sayıda güvenlik açığı var.” şeklinde sınav güvenliğinin güvenilirliği tehdit ettiğini belirtmişlerdir.

Güvenirliğin sebeplerine ilişkin görüşler bildiren akademisyenlerin bu görüşlerinde güvenilirlik için alınabilecek önlemlere de değindikleri görülmüştür. Güvenirlik için alınabilecek önlem kategorisine ilişkin frekanslar Tablo 6’da verilmiştir.

Tablo 6. Güvenirlik için alınabilecek önlemler kategorisine ve alt kategorilerine ilişkin frekanslar

Güvenirlik için alınabilecek önlemler	Yüz yüze	Uzaktan
	Frekans	Frekans
Soru sayısının yükseltilmesi	2	2
Öğrencinin takıldığı yeri sorabilmesini sağlama	1	0
Sınav güvenliği için önlemler alma	0	1
Farklı ölçme araçları kullanma	0	1
Madde havuzu oluşturma	0	1
Süre sınırı koyma	0	1
Ezbere dayalı sorular sormama	0	1

Tablo 6 incelendiğinde yüz yüze sınavlarda akademisyenlerin daha güvenilir ölçümler elde edebilmek amacıyla en fazla (f=2) soru sayısının yükseltilmesinin önerildiği görülmüştür. Örneğin A10: “soru sayısının fazlalığı ile güvenilirliği yüksek sonuçlara ulaşmak mümkün.” şeklinde bu görüşünü ifade etmiştir. Ayrıca öğrencinin takıldığı yeri sorabilmesinin güvenilirliği artıracaklarını belirten bir akademisyen (A9) bulunmaktadır. Yüz yüze sınavlara benzer olarak uzaktan sınavlarda da akademisyenlerin daha güvenilir ölçümler elde edebilmek amacıyla alınabilecek önlemler arasında en fazla (f=2) soru sayısının yükseltilmesini belirttikleri görülmüştür. Örneğin soru sayısının arttırılması önlemini alan A13 şu şekilde ifade etmiştir:

Uzaktan eğitim sürecinde bilgisayar ortamında sınava girdiklerini düşünürsek eş zamanlı kopya çekmeleri ve birbirlerine yanıtlarını paylaşıp cevaplarını değiştirmelerinin önüne geçmek adına güvenilirlik önlemleri almaya çalıştım. Süreçte kopya çekmenin önüne geçmek için sorulara geri dönüşü engelleyip eş zamanlı kamera açmalarını istedim. Bu sadece sınav ortamının güvenilirliği adına aldığım bir önlemdi. Niteliksel olarak güvenilirliği sağlamak için ise her konudan çok soru sormaya çalışarak duyarlılığı arttırmaya çalıştım.

Bu yanıtında A13 uzaktan sınavlarda alınabilecek başka önlemlere de değinmiştir ve bunlar sınav güvenliği için önlemler alma alt kategorisini oluşturmuştur. Uzaktan sınavlarda daha güvenilir ölçümlerin elde edilmesi için alınabilecek diğer önlemler farklı ölçme araçları kullanarak ölçümleri çeşitlendirmek, madde havuzu oluşturmak, sınav süresini sınırlandırmak, cevabın kaynaklarda net biçimde olmadığı sorular yazmak gibi önlemlerdir.

### iii. Kapsam Geçerliği

Kapsam geçerliğinin mevcut durumu kategorisine ilişkin alt kategoriler kapsam geçerliğine kanıt sunulmuş / yüksek ve kapsam geçerliği düşük olarak belirlenmiştir. Bu alt kategorilere ait frekanslar Tablo 7’de sunulmuştur.

Tablo 7. *Kapsam geçerliğinin mevcut durumu kategorisine ve alt kategorilerine ilişkin frekanslar*

Kapsam geçerliğinin mevcut durumu	Yüz yüze	Uzaktan
	Frekans	Frekans
Kapsam geçerliğine kanıt sunulmuş / yüksek	15	11
Kapsam geçerliği düşük	1	4

Tablo 7 incelendiğinde hem yüz yüze sınavlarda hem de uzaktan sınavlarda kapsam geçerliğinin genel olarak yüksek olduğu akademisyenler tarafından belirtilmiştir. Ancak uzaktan sınavlarda yüz yüze sınavlara göre kapsam geçerliğinin düşük olduğunu belirten daha çok akademisyen bulunmaktadır. Kapsam geçerliğinin yüz yüze sınavlarda yüksek olduğunu belirten 15 akademisyen bulunmaktadır. Örneğin yüz yüze sınavlarda kapsam geçerliğinin yüksek olduğunu A3: “Kapsam geçerliği yüksek sınavlar yapabiliyorum” ve A10: “çok sayıda soru sorabiliyorum. Kapsam geçerliği açısından bir sıkıntı yok.” şeklinde ifade etmiştir. Uzaktan sınavlarda kapsam geçerliğinin yüksek olduğunu belirten 11 akademisyen bulunmaktadır. Bu akademisyenlerden bazılarının görüşleri şu şekildedir: “Kapsam Geçerliğim yüksektir.” (A4), “Ölçme ve değerlendirme sürecinin kapsam geçerliği son derece yüksektir” (A5), “Hem yüz yüze eğitimde hem de uzaktan eğitimde vizede işlenen konuların ağırlığını düşürerek finalde eksiksiz bir şekilde tüm konuları yoklamaya çalıştım.” (A13). Diğer yandan yüz yüze sınavlarda kapsam geçerliğinin düşük olduğunu belirten yalnızca bir akademisyen bulunurken, uzaktan sınavlarda dört akademisyen bulunmaktadır. Akademisyenlerden A7 yüz yüze sınavlar için “ders süresinden dolayı nispeten düşük oluyor.” şeklinde ifade etmiştir. Uzaktan sınavlar için ise A9: “kapsam geçerliği yüz yüzeyle kıyasla daha düşük olabilmektedir.” şeklinde görüşünü belirtmiştir. Kapsam geçerliğinin mevcut durumuna ilişkin görüşler bildiren akademisyenlerin bu görüşlerinde mevcut durumun nedenlerine de değindiği görülmüştür. Kapsam geçerliğinin mevcut durumunun nedenleri kategorisine ilişkin frekanslar Tablo 8’de verilmiştir.

Tablo 8. *Kapsam geçerliğinin mevcut durumunun nedenleri kategorisine ve alt kategorilerine ilişkin frekanslar*

Kapsam geçerliğinin mevcut durumunun nedenleri	Yüz yüze	Uzaktan
	Frekans	Frekans
<b>Kapsam geçerliğine kanıt sunulmuş/yüksek</b>		
Belirtke tablosu kullanma	4	3
Konuları iyi şekilde örnekleme	4	5
Kritik kazanımların yoklanması	1	0
Çok sayıda soru ve sınav yapma	1	1
<b>Kapsam geçerliği düşük</b>		
Ders süresi	1	1
Madde havuzu kullanılması	0	1
Sorulardaki öğelerin kısıtlılığı	0	1

Tablo 8 incelendiğinde yüz yüze sınavlarda kapsam geçerliğine kanıt sunulmuş/yüksek olmasının sebeplerinden en çok belirtke tablosu kullanmaya (f=4) ve konuları iyi şekilde örnelemeye (f=4) değinmişlerdir. Örneğin A11 bunu “Amaçlanan kapsamı ölçmek için belirtke tablosu kullanım ve hangi amaç için sınav yaptığımı belirlerim öncesinde...” şeklinde ifade etmiştir. Akademisyenlerden A15 ise “Belirtke tablosunu tüm öğrenme çıktılarını baz alarak hazırlıyorum ve tüm araçlar da buna hizmet ediyor.” şeklinde yaptığı uygulamaya değinmiştir. Konuları iyi şekilde örnelediğini belirten akademisyenlerden ise A6: “Dersi kapsamında anlattığım konuları ilgili konunun kaç hafta ve ne düzeyde ele alındığını belirtke tablosu olmasa da konu başlığı ve ele alınma süresi bağlamında ağırlıklandırarak sınav sorularına buna göre dağıtıyorum.” şeklinde ifade ederken, A14: “Çoktan seçmeli testlerde gösterdiğim her konudan soruyor, soruları eşit dağıtmaya çalışıyorum.” Şeklinde ifade etmiştir. Yüz yüze sınavlarda kapsam geçerliğinin düşük olma nedenlerine değinen bir akademisyen (A7) bulunmaktadır ve kapsam geçerliğinin ders süresinden dolayı nispeten düşük olduğunu belirtmiştir. Ayrıca kapsam geçerliğinin yüksek olduğunu ya da kanıt topladığını belirten akademisyenlerden A9 kritik kazanımları yokladığına, A10 ise çok sayıda soru sorduğunu ve birden fazla sınav yaptığını değinmiştir. Uzaktan sınavlarda da yüz yüze sınavlara benzer olarak kapsam geçerliğinin yüksek olmasının nedenlerinden en çok konuları iyi şekilde örnelemeye (f=5) değinmişlerdir. Bunu belirtke tablosu kullanma (f=3) takip etmiştir. Buna göre belirtke tablosunun kullanımının uzaktan sınavların geliştirilmesinde azaldığı söylenebilir. Konuları iyi şekilde örnelediğini ifade eden akademisyenlerden A4: “İşlemiş olduğum tüm konuları sınavda iyi bir şekilde örnelemeye çalışıyorum.” şeklinde bunu belirtirken, A14: “Ödevde soruların verdiğim konulara göre dağılımına oldukça dikkat ediyorum” şeklinde belirtmiştir. Ek olarak uzaktan sınavlarda da A10 yüz yüze sınavlarda olduğu gibi çok sayıda soru sorduğunu ve birden fazla sınav yaptığını belirtmiştir. Uzaktan sınavlarda kapsam geçerliğinin düşük olmasına üç farklı neden sunulmuştur. Bunlar ders süresinin kısıtlı olması, madde havuzu kullanıldığı için her öğrenciye farklı sınav uygulanması ve sorularda görsel, tablo, şekil gibi öğelerin uzaktan sınavlarda kullanılamamasıdır.

#### iv. Sınav Güvenliği

Akademisyenlerin görüşleri incelendiğinde sınav güvenliği temasına ilişkin iki kategorinin oluştuğu görülmüştür. Sınav güvenliği temasına ilişkin kategoriler mevcut güvenlik durumu ve güvenlik için alınan/alınabilecek önlemler olarak belirlenmiştir. Mevcut güvenlik durumu kategorisi ile ilişkili alt kategorilere ait frekanslar Tablo 9’da verilmiştir.

Tablo 9. Mevcut güvenlik durumu kategorisine ve alt kategorilere ilişkin frekanslar

Mevcut sınav güvenliği durumu	Yüz yüze	Uzaktan
	Frekans	Frekans
Sınıf ortamı	2	0
Akademisyenin sorumluluğu olmaması	4	2
Sınav güvenliği yüksek	1	0
Sınav güvenliği düşük	0	5
Sistemin güvenliği	0	1

Tablo 9 incelendiğinde yüz yüze sınavlarda sınav güvenliğinin mevcut durumuna ilişkin en çok (f=4) sınav güvenliğinin akademisyen sorumluluğunda olmadığına değinerek özellik sınav görevlilerinden bahsetmişlerdir. Bunu belirten akademisyenlerden A6: “maalesef bu konuda sınavda görevli asistanlardan belirlenmesi aşamasında yönetim sorumlu olduğu için bu konu benim dışımda biraz. Ben her ne kadar belli aralıklarla sınıfları dolaşsam da bu görev sınavda görevli asistanın sorumluluğunda olduğunu düşünüyorum.” şeklinde ifade ederken, A12: “Her sınav salonunda bir gözetmen yer alıyor ve sınav sorumlusu süreçte oluyor. Yani sınav güvenliği için mümkün tüm koşullar sağlanmaya çalışılıyor.” şeklinde ifade etmiştir. Mevcut sınav güvenliği durumu kategorisinde gözetmen sorumluluğunu, sınıf ortamı alt kategorisi (f=2) takip etmektedir. Bu alt kategoriye ilişkin akademisyen A2: “sınıflarımız çok uygun ortamlara sahip olduğundan kopya girişimi neredeyse sıfır.” şeklinde görüşünü bildirmiştir. Son olarak yüz yüze sınavlarda sınav güvenliğinin yüksek olduğunu ifade eden bir akademisyen bulunmaktadır. Uzaktan sınavlarda sınav güvenliğinin mevcut durumu incelendiğinde akademisyenlerin en çok (f=5) sınav güvenliğinin düşük olduğunu belirttikleri görülmektedir. Bu konuda akademisyenlerden, A16: “Çok sayıda güvenlik açığı var.” şeklinde görüş bildirirken, A17: “Uzaktan eğitimde sınavı sadece yapmış olmak için yapıyoruz. Gerçekten sınava çalışan öğrenciler de var ama maalesef kopya ile sınavını tamamlayan öğrenciler de var.....Ama yine de öğrencilerin kopya çektiklerini sınav sonuçlarından anlayabiliyorum.” şeklinde görüş bildirmiştir. Uzaktan sınavlar için sınav güvenliği düşük alt kategorisini yüz yüze sınavlar ile ortak bir alt kategori olan akademisyenin sorumluluğu olmaması takip etmektedir (f=2). Örneğin A3 bunu şu şekilde ifade etmiştir: “Güvenlik önlemlerini bireysel olarak alma şansım çok az”. Son olarak uzaktan sınavlarda sistemin güvenliği kadar sınavın güvenli olduğunu belirten bir akademisyen bulunmaktadır (A4). Sınav güvenliğinin mevcut durumunun yanı sıra sınav güvenliği için alınan ve alınabilecek önlemlere de akademisyenlerin değindiği görülmüştür. Güvenlik için alınan/alınabilecek önlemler kategorisine ilişkin frekanslar ise Tablo 10’da sunulmuştur.

Tablo 10. *Sınav güvenliği için alınan/alınabilecek önlemler kategorisine ve alt kategorilerine ilişkin frekanslar*

Sınav güvenliği için alınan/alınabilecek önlemler	Yüz yüze	Uzaktan
	Frekans	Frekans
<b>Güvenliği akademisyenin kendisinin sağlaması</b>		
Farklı formlar oluşturma	7	0
Sınavda bulunma	2	0
Psikolojik yönetim	2	1
Sınıf içi değerlendirmelerden yararlanma	1	1
Avantaj sağlayan materyalleri toplama	1	0
Süre sınırı koyma	0	2
Ezbere dayalı sorular sormama	0	1
Ödevi el ile yazdırma	0	1
<b>Sistemsel ve yönetsel önlemler alınması</b>		
Soruları/seçenekleri karıştırma	0	6
Sistemde madde havuzu oluşturulması	0	2
Kamera/ses açılmasını sağlama	0	2
Gözetmen sayısının artırılması	2	0
Soruları bir kez görmeyi sağlanması	0	1

Tablo 10 incelendiğinde yüz yüze sınavlarda sınav güvenliği için alınan ya da alınabilecek önlemlere yönelik görüşler incelendiğinde akademisyenlerin en çok (f=7) farklı formlar oluşturmayı belirttikleri görülmüştür. Örneğin A1: “En az iki grup yapıyorum, gruplarda hem soruların hem seçeneklerin yerini değiştiriyorum” şeklinde görüşünü bildirerek test formlarını çeşitlendirdiğini ve bunu yaparken hem soruların hem seçeneklerin yerlerini değiştirdiğini belirtmiştir. Diğer akademisyenlerden A7 “a ve b gruplarının oluşturulması sınavlar için” şeklinde öneri sunmuş, A10 ise “kalabalık gruplar olduğu için A ve B formları oluşturuyorum.” şeklinde kendi aldığı önleme değinmiştir. Güvenliği akademisyenin kendisinin sağlamasında yüz yüze sınavlarda farklı formlar oluşturmayı sınavda bulunma (sınavda akademisyenin bizzat kendisinin bulunması, f=2) ve psikolojik yönetim (f=2) takip etmektedir. Psikolojik yönetimde akademisyenler, öğrencilere dürüstlük beklentilerini vurguladıklarını ve onları kopya çekmemeye motive ettiklerine ilişkin görüşler bildirmişlerdir. Ayrıca sınıf içi değerlendirmeleri ağırlıklı olarak kullanma ve yazılı materyaller, cep telefonu, hesap makinesi gibi soruların yanıtlanmasında öğrencilere avantaj sağlayacak materyalleri toplama akademisyenlerin aldığı ya da alabileceği önlemler arasında yer almaktadır. Son olarak yüz yüze sınavlarda sistemsel ve yönetsel önlemlerin arasında gözetmen sayısının artırılması (f=2) bulunmaktadır. Akademisyenler fakülte yönetimlerinden gözetmen ataması yapılmasını/gözetmen sayılarının artırılmasını beklemektedir.

Sınav güvenliği için alınan/alınabilecek önlemlerde uzaktan sınavlara ilişkin bulgular incelendiğinde akademisyenlerin en çok (f=6) soruları/seçenekleri karıştırmayı ifade ettikleri görülmüştür. Örneğin A8: “kopyanın önüne geçmek için süre sınırı olan soruların random geldiği bir sınav tasarlıyorum, A17: “Önem olarak soruları karışık sunuyoruz cevapları karışık sunuyoruz.” ifadelerine görüşlerinde yer vermiştir. Sistemsel ve yönetsel olarak alınabilecek bu önlemlerde soru/seçeneklerin karıştırılmasının yanı sıra sistemde madde havuzunun sorulacak soru sayısından fazla girilmesini ve öğrencilerin kamera ve seslerinin açık olmasına değinmişlerdir. Uzaktan sınavlarda akademisyenlerin kendilerinin alabilecekleri önlemler incelendiğinde ise en çok (f=2) sorulara ve sınava süre sınırı koyduklarını belirtmişlerdir. Sonuç olarak akademisyenler yüz yüze sınavlarda sınav güvenliğini kendi sağlamaya yönelik daha fazla görüş bildirirken, öneri, uzaktan sınavlarda sistemsel ve yönetsel olarak alınabilecek önlemlere daha çok değinmişlerdir.

#### **v. Uygulama Türü**

Akademisyenlerin ölçme ve değerlendirme süreçlerinin yüz yüze ve uzaktan yürütülmesi tercihlerine ilişkin görüşleri bu alt problemde incelenmiştir. Araştırmaya katılan 17 akademisyenin 16'sının (%94,11) yüz yüze sınavları tercih ettiği bulunmuştur. Yalnızca bir akademisyen uzaktan sınavlarını tercih ettiğini belirtmiş ve bu tercihinin nedenlerinden şu şekilde bahsetmiştir: “Uzaktan tercih ederim. Sınavı okuma mizanpaj, fotokopi bir iş yükü olmaktan çıkıyor. Ayrıca kâğıttan tasarruf sağlanıyor ve sınav evraklarını saklama gibi bir sorun oluşmuyor.” (A1). A1 görüşlerinde genel olarak



uzaktan sınavların kullanılabilirliğine ve ekolojik duyarlılığa vurgu yapmıştır. Yüz yüze sınavları tercih eden akademisyenlerden bazıları da tercih etme nedenlerine değinmiştir. Örneğin; A4: “Çünkü daha güvenilir ve geçerli bir sınav yapılıyor.”, A6: “Kesinlikle yüz yüze olmasını tercih ederim. Uzaktan yapılan sınavlar.....kopya çekilme olasılığı çok yüksek olduğu için uzaktan sınavları önemsemiyorum.”, A9: “Yüz yüze sınavı kapsam geçerliği, test güvenliği bağlamında tercih ederim.”, A10:

Yüz yüze sınavı tercih ederim. Lisans düzeyinde öğrenciler hala ders geçme odaklılar. Bu sebeple online sistemlerde sınav yazılımının/sürecin açığını bulup mutlaka kopya çekmeye kalkıyorlar. Bu durumda sınavdan elde edilecek sonuçların güvenilirliğini ve geçerliğini olumsuz etkiliyor. Başarılı/başarısız tüm öğrencilerin geçmesi ile süreç sonuçlanıyor. Sınavlar online olunca derse olan ilgi azalıyor. Ayrıca ölçmede işlem yapmayı gerektirecek konularımızda oluyor. Bu tür işlemler soruların çözümü için de kâğıt kalem sınavları daha pratik.

ve A14: “Kesinlikle yüz yüze çünkü öğrenci daha ciddiye alıyor. Bir de bilmediği ve KPSS'de çıktığını görünce sınıf içinde çok motive oluyor. Bu durum bizi de motive ediyor.” şeklinde yüz yüze sınavları tercih etme nedenlerini belirtmişlerdir. Akademisyenler belirttikleri nedenler incelendiğinde genel olarak güvenilirlik, geçerlik, sınav güvenliği, kullanılabilirlik ve motivasyon açısından yüz yüze sınavların tercih edildiğini göstermektedir.

Son olarak yüz yüze sınavları tercih ettiğini belirten A4 ek açıklamalarda da bulunmuştur:

Aslında uzaktan öğretimin avantajları da var. Şahsi fikrim süreçte birlikte kullanılmaları. Özellikle alternatif değerlendirme yaklaşımı kullananlar, ya da ödev vs. için uzaktan eğitim sistemi çok kullanışlı. Bu kısımda uzaktan öğretimden yararlanılabilir. Ama sınav yapılacaksa bu mutlaka gözetmen eşliğinde, kalem ve kâğıtla yüz yüze olmalıdır.

Bu açıklamalarında A4 yüz yüze sınavları tercih etmesine karşın tamamlayıcı değerlendirmelerin uzaktan yürütülmesinin daha kullanışlı olduğunu ifade etmiştir.

### **Sonuç ve Tartışma**

Araştırmada ölçme ve değerlendirme süreçlerini yüz yüze ve uzaktan gerçekleştiren ölçme ve değerlendirme alanında çalışan akademisyenlerin her iki uygulamaya ilişkin görüşleri incelenmiştir. Her iki uygulama sürecinde çoktan seçmeli testlerin daha çok tercih edildiği belirlenmiştir. Çoktan seçmeli testlerin kullanılma nedenleri arasında ilk olarak kullanılabilirlik belirtilirken ikinci olarak kapsam geçerliği ve güvenilirliğinin yüksek olması belirtilmiştir. Çoktan seçmeli testlerin özellikle kalabalık gruplarda uygulama kullanılabilirliğinin her iki uygulama da etkili olduğu belirlenmiştir. Çoktan seçmeli madde ile fazla sayıda soru sorulabilmesi nedeniyle kapsam geçerliğinin yüksek olması, ayrıca objektif puanlanabilmesinden dolayı güvenilirliğinin yüksek olması, kısa zamanda kalabalık gruplara uygulanabilirliği bu madde türünün özelliklerindedir (Özçelik, 2010; Turgut, 1992). Ayrıca hemen hemen tüm eğitim düzeylerinde öğretmenler geleneksel ölçme araçlarına daha aşinadır ve en çok kullanılan geleneksel ölçme araçlarından birisi de çoktan seçmeli maddelerdir (Anıl ve Acar, 2008;

Çakan, 2004, Gelbal ve Kelecioğlu, 2007). Çoktan seçmeli testler ayrıca uzaktan ölçme ve değerlendirme süreçlerinde kullanılan öğrenme yönetim sisteminde de kolayca uygulanabilmektedir (Taşkıran, 2021). Akademisyenler ikinci sırada yüz yüze uygulamada karma testleri seçerken uzaktan yürütülen süreçlerde ise ödevleri tercih etmişlerdir. Yüz yüze süreçlerde karma testlerin tercih edilme nedenleri arasında farklı seviyede soruların sorulabilme imkânı sunması belirtilmiştir. Ölçme araçlarının farklı üstünlüklerinin mevcut olmasından dolayı farklı ölçme araçlarının yer aldığı karma testler tercih edilebilir. Yüz yüze uygulamalarda en az kısa cevaplı ve yazılı yoklamalar tercih edilmiştir. Yazılı yoklamanın tercih nedeni olarak çoktan seçmeli sınavın oluşturulamadığı durumlarda hazırlama kolaylığının olması belirtilmiştir. Uzaktan eğitim süreçlerinde ise en çok tercih edilen ikinci sıradaki ölçme aracı ödev olarak belirtilirken en az yazılı yoklama ve karma testler belirtilmiştir. Akademisyenler uzaktan ölçme ve değerlendirme ödevi öğrenci motivasyonunu sağlama ve sınav güvenliği açısından tercih ettiklerini belirtmiştir. Öğrenciler ödevler ile yeni bilgiler öğrendiklerini düşünmekte ve notlarını da iyileştirebileceklerini düşünerek motive olabilmektedir (Featherstone, 1985; akt: Kan, 2009). Karataş (2014) çalışmasında bu araştırmanın sonuçlarına benzer şekilde uzaktan eğitim öğrencilerinin en çok tercih ettikleri 11 ölçme aracını sıralamalarını istemiş, ödev ilk 5 araç içinde yer alırken yazılı yoklamalar son sırada yer almıştır. Usta ve diğerleri (2022) çalışmasında da uzaktan değerlendirilme süreçlerinde uzmanların ödev verilmesi yönünde görüşlerinin olduğunu belirtmiştir. Araştırmada ayrıca uzaktan süreçlerde yüz yüzeden farklı olarak ölçme araçlarının tercih edilme nedenleri arasında sınav güvenliği gösterilmiştir. Özellikle sınav güvenliği uzaktan ölçme ve değerlendirme süreçlerinde önemli risk konuları arasındadır (Acar Güvendir ve Özer Özkan, 2021; Atılğan ve diğerleri., 2020; Bozkurt ve Uçar, 2018; James, 2016; Shraim, 2018). Sınav güvenliğini sağlamaya yönelik literatürde farklı uygulamalar önerilmiştir (Al-Shalout, Rasmi ve Hassan, 2021; Bedford, Gregg ve Clinton, 2011; Bozkurt ve Uçar, 2018; James, 2016). Kimlik doğrulama sınıflandırmaları kullanılması, biyometri uygulamaları ile birleştirilmiş gözetmen eşliğinde veya elektronik gözetmen yazılımları eşliğinde uygulamalar gerçekleştirilebilir. Parmak okuyucular, 360 derece kameralar kullanılabilir. Sınav esnasında öğrenciye özel güvenlik soruları bir veya birden fazla sorulabilir. Ayrıca etik kuralların belirlenmesine yönelik çalışmalar da gerçekleştirilebilir (Bozkurt, 2020).

Araştırma kapsamında ölçme ve değerlendirmenin önemli konularından olan güvenilirlik ve geçerlik konuları da incelenmiştir. Akademisyenlerin çoğunluğu ölçme ve değerlendirme süreçlerinin yüz yüze gerçekleştirildiğinde yüksek güvenilirliğe sahip olduğunu belirtmişlerdir. Bunun tam tersi olarak uzaktan yürütülen süreçlerde ise ölçme ve değerlendirme süreçlerinin düşük güvenilirliğe sahip olduğunu belirtmişlerdir. Yüz yüze de yüksek güvenilirlik nedenleri arasında soru sayısı, grup heterojenliği, sınav güvenliğinin sağlanması, madde istatistikleri bilinen madde havuzunun kullanılması belirtilmiştir. Uzaktan eğitimde ise en çok sınav güvenliği endişesinden dolayı güvenilirliğin düşük olduğu belirtilmiştir. Güvenirlik temasında dikkat çekici konularından bir tanesi

teknoloji kullanımının güvenilirliği uzaktan süreçlerde olumsuz etkilediğidir. Sharim (2018) araştırmasında da bu araştırmanın sonuçlarına benzer şekilde katılımcıların %70,5'inin kullanılan teknolojiden kaynaklı testin güvenilirliğini düşük bulduklarını belirtmiştir. Teknoloji kullanımının sınav sürecinde hazırlama, uygulama, düzeltme gibi aşamalarında olumlu yansımaları (Hack ve Tarouco, 2000; akt: Karadağ, 2021) söz konusu olsa da birtakım endişeler de oluşturmaktadır. Burada teknolojik alt yapı, güvenlik sorunu, teknoloji kullanabilme yeterliği vb. bu sorunlar arasında olabilir. James (2016)'ın üniversite öğrencileri ile gerçekleştirdiği çalışmada, öğrenciler online sınavı tercih etmemesinde teknoloji ile ilgili sorun yaşama endişesinin söz konusu olduğunu belirtmiştir. Dolayısıyla teknoloji ile ilgili sorun yaşama endişesi de öğrencilerin motivasyonunu olumsuz etkileyebilecektir.

Testin güvenilirliği artırmak için alınabilecek önlemler arasında her iki uygulamada soru sayısının artırılması belirtilmiştir. Soru sayısının artırılması ile ölçme aracının duyarlılığını da artıracığından güvenilirliğe olumlu etkileyecektir. Usta ve diğerleri (2022)'nin çalışmasında uzmanların çevrim içi sınavlarda ölçme ve değerlendirme temasına ilişkin soru sayısının artırılması, ölçme araçlarında çeşitliliğin sağlanması yönünde öneriler sunduğunu belirtmiştir. Araştırma kapsamında yüz yüze de uzaktan süreçlerden farklı olarak öğrencinin anlamadığı hususu sorabilmesinin güvenilirliğe olumlu etkisinin olabileceği belirtilmiştir. Uzaktan ölçme ve değerlendirme süreçlerinde ise sınav güvenliğini sağlamaya yönelik alınacak önlemler, farklı ölçme araçlarının kullanılması, ezber dayalı sorular sormama, madde havuzu oluşturma gibi unsurların güvenilirliğe olumlu katkısı olacağı belirtilmiştir. Chaudhary ve Dey (2013), açık ve uzaktan eğitimde değerlendirme sürecini inceledikleri çalışmalarında değerlendirmenin çeşitli araçlar yoluyla gerçekleştirilmesini belirtmişlerdir. Yazılı sınavlar, kısa yanıt ve objektif testler, sözlü sınavlar, uygulamalı sınavlar, alan temelli uygulama ve değerlendirme, proje ve performans değerlendirme gibi araçları önermişlerdir. Yağcı, Ekiz ve Gelbal (2015) çalışmalarında da birden fazla doğru yanıt olan çoktan seçmeli madde kullanımının da güvenilirliği artırabileceğini belirtmişlerdir. Araştırmanın bulguları genel olarak değerlendirildiğinde uzaktan yürütülen ölçme ve değerlendirme süreçlerinin ölçme hatalarına daha açık olduğu söylenebilir.

Yüz yüze ve uzaktan süreçlerde kapsam geçerliğinin incelenmesinde ise hem yüz yüze hem de uzaktan ölçme ve değerlendirme süreçlerinde kapsam geçerliğinin sağlandığına ilişkin kanıt sunulmuş ve yüksek olduğu belirtilmiştir. Ancak uzaktan ölçme ve değerlendirme süreçlerinde kapsam geçerliği düşük olduğunu belirtenlerin sayısı yüz yüze süreçlerine kıyasla daha yüksektir. Maddelerin konuları örneklemesi her iki uygulamada kapsam geçerliğinin yüksek olma nedenleri arasındadır. Araştırmada uzaktan yürütülen süreçlerde sınav hazırlamada maddelerin içeriğindeki grafik, şekil, tablo gibi unsurların sisteme yüklenememesindeki kısıtlılıkların daha düşük kapsam geçerliğine neden olduğu da belirtilmiştir. Bu durumun uzaktan eğitimde kullanılan sistemden kaynaklı bir sınırlılık olduğu düşünülmektedir. Ancak genel olarak değerlendirildiğinde her iki uygulamada da kapsam geçerliğinin sağlanabildiği sonucu çıkarılabilmektedir.

Araştırma kapsamında ele alınan bir diğer tema da güvenlik olup her iki uygulama için akademisyenlerin görüşleri ele alınmıştır. Akademisyenlerin çoğunluğu yüz yüze sınavlarda gözetmenlerin görevli olmasından dolayı sorumluluğunun olmadığını belirtmişlerdir. Uzaktan ölçme ve değerlendirme süreçlerinde ise en çok sınav güvenliğinin düşük olduğu belirtilmiştir. Ayrıca yüz yüze uygulamalarda sınav güvenliğini akademisyenin kendisinin sağlayabileceği önlemler belirtmişlerdir.

Al-Shalout ve diğerleri (2021) çalışmalarında uzaktan yürütülen sınavların güvenliğini sağlamaya yönelik teknikleri incelemiştir. Tüm güvenliği artırmaya dönük alınabilecek önlemlere rağmen, elektronik ortamda yürütülen uygulamaların okulda veya üniversite salonlarında gerçekleşen sınavlar ile rekabet edemeyeceğini de belirtmişlerdir. Araştırmada güvenlik temasında dikkat çeken bir diğer bulgu ise farklı formların oluşturulmasıdır. Özellikle kopya çekme davranışının önüne geçilmesinde, kalabalık sınıflarda farklı formların oluşturulması yoluna gidilmesi tercih edilebilir. Farklı formların oluşturulmasında içerik dengelemesinin sağlanması gereklidir. Ancak madde güçlüğü, madde ayırt ediciliğinin denk olarak oluşturulmadığı durumlarda da soruların yerlerinin değiştirilmesine dayalı olarak oluşturulan formlar da kullanılabilir. Çoğunlukla soruların yerlerinin değiştirildiği farklı formların kullanılması tercih edilmektedir. Yağcı ve diğerleri (2015) çalışmalarında her öğrenciye farklı soruların sorulmasını ya da aynı soruların öğrencilere farklı doğru yanıtlar ile sunulması test güvenliğini artırmak amaçlı kullanılabileceğini de belirtmişlerdir. Araştırmada yüz yüze uygulamada ise çoğunluk sistemselsel ve yönetselsel önlemlerin gerekliliğini belirtmişlerdir. Whitley (1998) lise öğrencilerinin kopya çekme davranışına ilişkin gerçekleştirdiği araştırmasında değerlendirmelerde daha çok mutlak değerlendirmenin tercih edilmesini, küçük sınıflarda öğrencilerin geniş bir alanda oturmasını eğer mümkün değil ise farklı test formlarının oluşturulması ve öğrencilerin dikkatle sınav boyunca izlenmesini kopya çekme davranışına yönelik önlem olarak sunmuştur.

Araştırma kapsamında son olarak akademisyenlere yüz yüze ve uzaktan ölçme ve değerlendirme süreçlerine ilişkin tercihleri sorulmuştur. Akademisyenlerde sadece bir kişi hariç yüz yüze uygulamayı tercih ettiğini belirtmişlerdir. Yüz yüze süreçlerin tercih edilmesinde ölçme araçlarının geçerlik, güvenilirlik ve kullanılabilirliğinin sağlanması açısından tercih edildiği belirtilmiştir. Genel olarak değerlendirildiğinde her iki ölçme ve değerlendirme süreçlerinin birbirlerine kıyasla üstünlük ve sınırlıkları olduğu görülmektedir. Ancak ölçme araçlarının psikometrik özellikleri açısından yüz yüze ölçme ve değerlendirme süreçlerinin daha tercih edilebilirliği üzerinde de durulmuştur. Araştırmanın sonuçlarından yola çıkılarak deneysel araştırmalar gerçekleştirilerek her iki uygulamaya ilişkin sonuçların karşılaştırılması araştırmacılara önerilebilir. Araştırma kapsamında ele alınan temalardan farklı temaların da ele alınarak süreçlerin karşılaştırılabileceği çalışmalar gerçekleştirilebilir.



## ENGLISH VERSION

### Introduction

With the rapid development of technology and the information age and its integration into education, the importance and usefulness of open and distance education are increasing day by day. In our country and the world, especially during the COVID-19 pandemic, distance education has rapidly spread to all education levels under emergency distance education, such as pre-school, primary, secondary, undergraduate, and graduate education. In this period, the teaching dimension of distance education was carried out in universities in Turkey through synchronous and asynchronous courses. As it can be understood from this point, in addition to the development of information technology, changes in social life have also been influential in the spread of distance education. In the distance education period, it is also essential to know how the teaching will be performed, how feedback will be given to the students, and how the success levels of the students will be determined. Because the finite interaction with students can be carried out within the framework of a certain perception, situations may arise in which the applications lose their depth or the system's applicability is limited (Karataş, 2003).

Despite the superiority of spatial independence of distance education (Sarı, 2020), measurement and evaluation processes also have significant limitations. The measurement and evaluation process is a feedback mechanism for the education system. Therefore, planning, examining, and organizing measurement and evaluation processes, which play an essential role in providing feedback to the system's administrators, teachers, students, parents, and other stakeholders, are always important issues in education. Especially in measurement and evaluation processes carried out distantly, feedback on the process, instant feedback given to the students, end-of-term exams, homework, and exam security are the issues emphasized (Acar Güvendir and Özer Özkan, 2021).

Assessment and evaluation processes are carried out for three purposes: diagnostic, formative, and summative, regardless of whether they are performed face-to-face or distantly (Özçelik, 2010). In diagnostic assessment, the inputs in the education system are evaluated before the teaching activities begin. On the other hand, formative assessment is conducted throughout the teaching process and aims to monitor development. Formative assessment has an essential place in the distance education process and face-to-face education because formative assessment aims to identify and eliminate learning

deficiencies. In addition, feedback is received on the effectiveness of the teaching process, and necessary improvements can be made. The summative assessment is carried out at the end of the teaching process. In this type of assessment, product/outcome evaluation is performed. Among the most common assessment types preferred in distance education are formative and summative assessment (James, 2016; Karadağ, 2014; Usta, Kılınc, and Okur, 2022; Shraim, 2019). Formative assessment is also critical in distance education because for the effectiveness of the assessment, both the student and the teacher should be active in the process, the teacher should regularly monitor the students' progress, and the students should be aware of their knowledge, skills, and abilities (Ocak and Karakuş, 2021).

In measurement and evaluation processes, students' knowledge, skills, and abilities are taken into account in the context of their cognitive skills. While classical assessment methods are mostly used to assess knowledge and skills, which are referred to as achievement, performance-based assessment methods are preferred for assessing abilities. Measurement and assessment tools used in the performance-based assessment are performance tasks, projects and portfolios, analytical and holistic rubrics, rating scales, and checklists (Gültekin, 2017; Kutlu, Doğan, and Karakaya, 2010; Popham, 1999; Şahin, 2019). There are different classifications in the use of measurement and evaluation tools, in addition to classical methods and performance-based assessments (Başol, 2018; Bıkmaz Bilgen, 2019; Doğan, 2009; Doğan, 2019; Kilmen, 2017; Özçelik, 2010). The first classification is traditional methods and complementary methods. Complementary measurement and evaluation tools are different from traditional tools and are included in assessment processes as a supplement to the limitations of traditional tools. Traditional measurement and evaluation tools can be presented as multiple-choice, open-ended, short-answer, matching, and true-false type items, oral exams, and written exams. Another classification used in categorizing measurement tools is objective and subjective measurement tools in terms of the error caused by the rater. While multiple-choice, true-false, and matching tests are objective, oral and written exams can be considered subjective. The existing classifications in the literature are categorized by considering the characteristics measured by the measurement tools, preparation, application, evaluation processes, and the steps used in the evaluation. Reliability, validity, and usefulness are the most important features that should be taken into consideration regardless of the classification of measurement tools. Validity is the most crucial feature that should be considered in terms of the level of measuring the feature that the measurement tool aims to measure. However, reliability, which is related to the accuracy of the scores obtained from the measurement tools and the accuracy of the decisions made, is another feature that needs to be investigated. After ensuring the validity and reliability of the measurement tool, another feature that should be considered is usefulness. The usefulness of the measurement tools should be evaluated by considering their economy, preparation-application time, ease of application-scoring-interpretation, and qualifications of the exam setter, practitioner, and respondent (Turgut, 1992).

Measurement tools provide a wealth of information to assessors in many ways but differ in terms of what they measure, reliability, validity, and usefulness. Considering all these criteria, it is very difficult to determine the best measurement tool. For this reason, regardless of whether the measurement and evaluation processes are conducted face-to-face or distantly, the reliability, validity, and usefulness of the processes should be ensured at a certain level because the examination of face-to-face and distance measurement and evaluation processes and suggestions for improvement are also critical in terms of the decisions to be made as a result of the evaluation.

There are various studies in the literature on the assessment of measurement and evaluation processes in the classroom and online/offline distance education. Acar Güvendir and Özer Özkan (2021) examined the evaluation processes in online exams compared to in-class exams based on pre-service teachers' opinions. Baran (2020) introduced the measurement tools used in open and distance education and examined different universities' measurement and evaluation approaches. Bozkurt and Uçar (2018) determined the requirements for security by taking the opinions of undergraduate students to ensure security in online measurement and evaluation processes. Karadağ (2021) discussed the preparation process for online exams and made suggestions. Ocak and Karakuş (2022) developed an attitude scale that determines university students' attitudes toward online exams based on technical elements, precautions, individual characteristics, and exam structure. Sarı (2020) discussed the measurement and evaluation process in distance education. Shraim (2019) examined university students' views on online exams. Taşkıran (2021) discussed the dimensions of management, learning, technology, and evaluation in distance education. Usta, Kılınç, and Okur (2022) conducted focus group interviews with experts working in the field of measurement and evaluation or open and distance learning at the universities and examined the opinions on the elements that should be considered in the measurement and evaluation processes of the learning environment in open and distance learning. Unlike the existing studies in the literature, this study aims to reveal the opinions of academicians working in the field of measurement and evaluation and conducting measurement and evaluation processes both face-to-face and distantly. In this direction, the views of academicians working in the field of measurement and evaluation were taken for both types of applications. Since they are experts in the field of measurement and evaluation, it was thought that they could examine the problems addressed in the research in depth in line with their areas of expertise. The research is also essential because it will provide information to practitioners to improve the reliability, validity, usefulness, and exam security of the measurement tools used in monitoring and evaluating students' achievements based on the research results. The research's sub-problems are presented below in line with the research's purpose.

1. Which assessment tools do academicians prefer in face-to-face and distance measurement and evaluation processes?

2. What are the opinions of academicians about the reliability of face-to-face and distance measurement and evaluation processes?

3. What are the opinions of academicians about the content validity of face-to-face and distance measurement and evaluation processes?

4. What are the opinions of academicians about the exam security of face-to-face and distance measurement and evaluation processes?

5. What are the preferences of academicians regarding the type of application in measurement and evaluation processes?

## Method

### Research Type

This study aims to examine both practices in the context of the characteristics determined within the scope of the research according to the opinions of academicians in the field of measurement and evaluation who perform both face-to-face and distance education. Since the study aims to provide rich definitions of an existing situation/phenomenon, it can be characterized as basic qualitative research (Meriam, 2009).

### Research Group

Within the scope of the research, 17 academicians working in the departments of measurement and evaluation of the faculties of education of universities and conducting measurement and evaluation processes for the "Measurement and Evaluation Course", both face-to-face and distantly, were reached and coded as A1, A2, A3, ..... A17. Demographic information about the academicians is presented in Table 1.

Table 1. *Demographic characteristics of the participants*

Variables		Frequency
Gender	Female	11
	Male	6
Title	Prof. Dr.	2
	Assoc. Dr.	8
	Assist. Dr.	5
	Res. Assist. Dr.	2
Years of teaching	2-5	6
	6-10	5
	11-15	4
	16 and above	2
Face-to-face teaching semester	1-5	6
	6-10	6
	11-15	3
Distance teaching semester	16 and above	2
	1-5	15
	6-10	2



Table 1 shows that most academicians in the study group consist of women (f=11). When the participants' titles and years of teaching are analyzed, it can be stated that they have different characteristics: the majority of the participants have the title of associate professor (f=8) and have 2-5 years of teaching experience (f=6). In terms of semesters in which the participants in the study group gave courses in face-to-face and distance education, the majority of the participants taught for one- five and six-ten semesters (f=5) in face-to-face education, and the majority of the participants taught for one- five semesters (f=15) in distance education.

### **Data Collection Tools**

Within the scope of the research, "Opinion Form on Face-to-Face and Distance Measurement and Evaluation Processes" was developed by the researchers. While developing the form, a draft form was created by interviewing three researchers giving the measurement and evaluation course and reviewing the literature. The draft form was finalized by taking the opinions of six measurement and evaluation experts. At this stage, the form was organized by evaluating each of the opinions on the form, presenting each new version to another expert, and following the process in this order. The final version of the form consists of two sections. The first part includes descriptive variables such as gender, title, seniority, and number of face-to-face and distance measurement and evaluation courses semesters. The second section contains five open-ended questions. The five open-ended questions reflect the problems of the research and consist of questions about the measurement tools, reliability, validity, exam security, and preferred application type that academicians working in the field of measurement and evaluation prefer in face-to-face and distance measurement and evaluation processes.

### **Data Analysis**

The research data were collected through Google Forms with the "Opinion Form on Face-to-Face and Distance Measurement and Evaluation Processes". The items in the data collection tool were determined to be compulsory to answer. The data obtained from academicians in the field of measurement and evaluation with this form were analyzed with content analysis. Content analysis was conducted by both researchers. As a result of content analysis, codes, subcategories, categories, and themes were created (Yıldırım and Şimşek, 2011). The themes formed in the content analysis were formed in parallel with the basic problems based on the questions in the form and the sub-problems of the research. Therefore, the findings are presented in the context of these themes.

### **Credibility, Transferability, and Consistency**

Direct quotations related to the categories and subcategories determined within the scope of the research were included in the reporting. This can be presented as evidence of credibility. On the other hand, the academicians included in the study group were purposively selected from the field of measurement and evaluation because the concepts addressed in this study, such as measurement and evaluation processes, measurement tools, validity, reliability, and exam security, are among the study

subjects of the field of measurement and evaluation. In addition, academicians from different academic titles with various teaching experiences were selected to form the study group. These factors can be presented as evidence of transferability (Arslan, 2022; Yıldırım and Şimşek, 2011).

In the study, the researchers coded separately within the scope of content analysis. Then, the percentage of agreement between these coders was calculated according to Miles and Huberman's (1994) percentage of agreement. The percentage of agreement between coders was found to be 91.18%. The researchers evaluated together the codings that caused inconsistency, and a consensus was reached in the units where there was a conflict, and thus the conflicts were resolved. In this direction, it can be said that evidence of consistency was provided within the scope of the research.

### Ethical Permissions of the Study

In this study, all rules specified by the "Higher Education Institutions Scientific Research and Publication Ethics Directive" were followed. None of the actions mentioned under the title of "Actions Contrary to Scientific Research and Publication Ethics", which is the second part of the directive, were carried out.

**The ethics board approval details:** Name of the ethics board = Gazi University Ethics Commission

Date of the decision = 19.12.2023

Document number = E.830002

### Findings

As a result of the content analysis conducted in the study, five themes were formed in parallel with the research problems. These themes are "measurement tools", "reliability", "content validity", "exam security", and "type of application". The findings of the study are presented in the light of these themes.

#### i. Measurement Tools

The frequencies of the measurement tools preferred by academicians in face-to-face and distance exams are presented in Table 2.

Table 2. *Measurement tools preferred by academicians*

Measurement tool preferred	Face-to-face	Distance
	Frequency	Frequency
Multiple-choice test	14	14
Short-answer test	1	0
Written examination	1	2
Mixed test	5	2
Homework	3	4

When Table 2 is examined, it is seen that academics mostly prefer multiple-choice tests ( $f = 14$ ) in face-to-face exams. When Table 2 is examined, it is seen that academics mostly prefer multiple-choice

tests ( $f = 14$ ) in face-to-face exams. For example, they preferred multiple-choice tests such as A2 stated, "I usually prefer to use multiple-choice tests for midterms and finals," and A9 said, "I do multiple-choice midterm and final exams." After multiple-choice tests, it was observed that mixed tests ( $f=5$ ) were preferred the most. For example, A15 stated that he preferred mixed tests, saying, "I want to determine the learning levels with tools containing different types of items." Short answer tests ( $f=1$ ) and written examinations ( $f=1$ ) were the least preferred measurement tools.

When the measurement tools preferred by academicians in distance exams are analyzed, it is seen that multiple-choice tests ( $f=14$ ) are preferred in distance exams as in face-to-face exams. For example, A4 stated to use both multiple-choice tests and written examinations:

The same is true for distance education. I generally prefer multiple-choice tests for midterm and final exams and classical tests for make-up exams. I prefer multiple-choice tests because they are useful, and I can easily score many questions. Since preparing a different test for the make-up exams is difficult, I prefer the classical exam.

As seen from the expression, he specified that he used similar item types in face-to-face and distance education. Similarly, A10 stated that he used multiple-choice tests in distance exams as "I used tests consisting of multiple-choice items in all my exams". It was observed that homework ( $f=4$ ) was the most preferred assessment tool after multiple-choice tests in distance exams, and this assessment tool was preferred more in distance exams than in face-to-face exams. A17, who used both multiple-choice tests and homework, stated that the measurement tool did not change according to whether it was administered distantly or face-to-face:

As in distance education, I made an exam consisting of multiple-choice items in the midterm and final exams, and I also asked them to write questions in the homework in accordance with the outcomes. In terms of assessment tools, distance or face-to-face did not differentiate the assessment tool I preferred.

When the findings related to mixed tests were analyzed, it was determined that they were preferred less than face-to-face education ( $f=2$ ). A12, who applied mixed tests, said, "In distance courses, I make quizzes at the beginning of the course to ensure that students study the subjects regularly. This method works because students do not actively participate in distance courses. I include different types of items in these quizzes" , and he stated that he diversified the types of items in the quizzes he conducted. In addition, although it was seen that written examinations were preferred more in distance exams than face-to-face exams ( $f=2$ ), it was observed that short-answer tests were not preferred by academicians in distance exams.

The categories, subcategories, and frequencies related to why academicians prefer these measurement tools in face-to-face and distance exams are given in Table 3.

Table 3. Frequencies related to the category and subcategories of the reasons why academicians prefer the measurement tools they prefer

Reasons for preference	Face-to-face	Distance
Multiple-choice test	Frequency	Frequency
Usefulness	8	7
Use of this test type in KPSS (Public Personnel Selection Examination)	1	1
Content validity and reliability	3	3
Objective scoring	2	2
Assessing learning and giving feedback	2	0
Exam security	0	1
Written examination		
Usefulness	1	1
Exam security	0	1
Karma test		
Diversity of subject areas and cognitive levels	1	0
Individual differences	1	1
Ensuring student motivation	0	1
Exam security	0	1
Different superiorities of item types	1	0
Homework		
Usefulness	1	0
Ensuring student motivation	0	1
Exam security	0	1

When Table 3 is examined, it is seen that academicians prefer multiple-choice tests in face-to-face exams primarily because of their usefulness ( $f=8$ ). A2 stated, "I also prefer multiple-choice tests because I work in a small university, and my course load is usually 30 hours or more". A4 stated, "I prefer multiple-choice tests because they are useful, and I can easily score a large number of questions". A10 said, "The number of students was always very high (about 300-350 students)". While three academics did not mention the reasons for preferring multiple-choice tests, three academics associate the reason for using multiple-choice tests with content validity and reliability. While A6, one of the academicians who stated that he preferred multiple-choice tests for content validity and reliability, expressed this as "the reason for using multiple-choice tests is that reliability and validity are high and they are scored objectively, as well". A7 expressed it as "for covering the outcomes to be measured to a significant extent and to make objective evaluation". A6 and A7 also emphasized the objective scoring in multiple-choice tests. It was observed that the only academician (A4) who used written examinations in face-to-face exams preferred this measurement tool due to its usefulness. On the other hand, it was observed that the three academicians who stated their reasons for preferring mixed tests had different reasons. These reasons can be listed as the fact that mixed tests cover every subject area and cognitive level and consider individual differences, and each item type has different advantages and limitations.

When the reasons for the preference of measurement tools in distance exams were analyzed, it was seen that, similar to face-to-face exams, multiple-choice tests were mostly preferred due to their

usefulness ( $f=7$ ). For example, A1 and A11 mentioned the crowded groups to which the exam will be applied by stating, "ease of scoring in large groups" and "because I teach a large number of students (at least 250 or more), respectively. Again, similar to face-to-face exams, it was determined that the category of usefulness was followed by content validity and reliability, which were among the reasons for preferring multiple-choice tests. Academicians who preferred multiple-choice tests for content validity and reliability in face-to-face and distance exams were the same participants (A5, A6, and A7). A striking finding is that an academician (A1) preferred multiple-choice tests in distance exams due to exam security and stated: "The ability to change the questions and options reduces security concerns." Although two academicians prefer written examinations, one of these academics (A4) stated that he used it in make-up exams due to its usefulness in the preparation of the written examination. At the same time, the other one (A5) emphasized exam security by stating that he preferred this measurement tool to minimize cheating in a limited time. On the other hand, although two academicians preferred mixed tests, these academics provided three different reasons. These reasons can be listed as providing student motivation, considering individual differences, and the fact that each item type has distinct advantages and limitations. Finally, two academicians (A7 and A14) stated that homework was one of the most preferred measurement tools after multiple choice in distance exams. It was determined that the reasons for these academicians to prefer homework were the same as those for preferring mixed tests and that it was used for student motivation and exam security reasons.

## ii. Reliability

The frequencies of academicians' views on the current reliability status are presented in Table 4.

Table 4. *Frequencies related to the current reliability status category and subcategories*

Current Reliability Status	Face-to-face	Distance
	Frequency	Frequency
High reliability	13	3
Low reliability	2	13

When Table 4 is examined, the majority of academics ( $f=13$ ) stated that face-to-face exams had high reliability. For example, A2 stated that the reliability was high: "When I do reliability analyses, it is generally high", A5: "The reliability of the measurements obtained from face-to-face tests is high", and A12: "It is quite high". However, in contrast to these, two academicians reported low reliability. For example, A7 stated that "reliability is low, but it is not bad in general" and that face-to-face exams had lower reliability than distance exams. When the findings related to distance exams were analyzed, most academicians ( $f=13$ ) stated that distance exams had low reliability. For example, A3: "There are serious reliability and security problems", A6: "I think that distance exam is not reliable", and A10: "Reliability is under threat." On the other hand, there are two academicians who stated that reliability was high in distance exams. For example, A15, who prefers mixed tests, stated: "I think that scoring reliability is high and bias is low."

Subcategories related to the reasons for the current reliability status were determined separately for high and low-reliability categories, and the frequencies of the subcategories are presented in Table 5.

Table 5. Frequencies for the category of reasons for the current reliability status and its subcategories

Reasons for the current reliability status	Face-to-face	Distance
	Frequency	Frequency
<b>High reliability</b>		
Exam security	1	0
Reduction of error sources	1	0
Number of items	1	0
Use of items whose items statistics are known	1	0
Heterogeneity of the group	1	0
<b>Low Reliability</b>		
Exam security	2	9
Number of items	0	1
Crowded classes	2	0
Technology	0	1
Guessing	1	0
Homogeneity of the group	0	1

When Table 5 is analyzed, it is seen that five academicians give reasons for high reliability in face-to-face exams, and each mentions different reasons. These reasons were exam security, reduction of error sources, number of items, use of items whose item statistics are known, and heterogeneity of the group. While one of the academicians (A5) stated that the reliability was high because the exam security was high: "Because cheating can be prevented", another academician (A6) noted that he minimized random errors because he knew the sources of error; therefore, his face-to-face exams were reliable: "As a matter of fact, I think I minimized possible random errors because I know the sources of error". It was observed that none of the academicians who stated that the reliability of the measurements in distance exams was high did not specify any reason. It was seen that the reasons for the decrease in reliability in face-to-face exams were collected in the subcategories of exam security, crowded classes, guessing, and homogeneity of groups. Exam security ( $f=2$ ) and crowded classes ( $f=2$ ) were prominent among these subcategories. For example, A7 emphasized both subcategories by saying, "due to the possibility of problems such as cheating in crowded classes." It was seen that the reasons for low reliability in distance exams were divided into subcategories, such as exam security, number of questions, technology, and homogeneity of the group. When the distribution of these subcategories was examined, it was determined that the academicians mostly specified exam security ( $f=9$ ) as the reason. For example, they stated that exam security threatened reliability: A1: "There are security concerns. Did he/she answer the questions himself/herself, did he/she get help, or did he/she use the Internet during the exam", A10: "Due to reasons such as exam security/cheating" and "It cannot be controlled where and how the students take the exam. Someone may be with them. Students may be together. They can benefit from print resources. There are many security flaws". They stated that exam security threatens reliability.

It was seen that the academicians who expressed opinions on the reliability reasons also mentioned the precautions that can be taken for reliability in these opinions. The frequencies related to the category of precautions that can be taken for reliability are given in Table 6.

Table 6. *Frequencies related to the category of precautions that can be taken for reliability and its subcategories*

Precautions that can be taken for reliability	Face-to-face	Distance
	Frequency	Frequency
Increasing the number of items	2	2
Allowing the students to ask where they do not understand	1	0
Taking precautions for exam security	0	1
Using different measurement tools	0	1
Creating an item pool	0	1
Setting a time limit	0	1
Avoid using questions that require memorization	0	1

When Table 6 is examined, it is seen that academicians mostly suggest increasing the number of items ( $f=2$ ) to obtain more reliable measurements in face-to-face exams. For example, A10 said, "It is possible to achieve high-reliability results with more questions". In addition, an academician (A9) stated that allowing the students to ask where they do not understand would increase reliability. Similar to face-to-face exams, it was observed in distance exams that among the precautions that can be taken to obtain more reliable measurements, academics mostly stated that the number of items ( $f=2$ ) should be increased:

Considering that they took the exam in a computer environment during the distance education process, I tried to take reliability precautions to prevent them from cheating simultaneously and changing their responses by sharing their answers. To prevent cheating, I prevented them from returning to the questions and asked them to turn on the camera simultaneously. Those were the precautions I took for the reliability of the exam environment. To ensure qualitative reliability, I tried to increase sensitivity by asking more questions from each subject.

In his response, A13 also mentioned other precautions that could be taken in distance exams, forming the subcategory of exam security measures. Other precautions that can be taken to obtain more reliable measurements in distance exams include diversifying the measurements by using different measurement tools, creating an item pool, limiting the exam time, and writing questions where the answer is not clear in the sources.

### iii. Content Validity

The subcategories related to the current status of the content validity category were determined as "evidence of content validity provided / high content validity" and "low content validity." The frequencies of these subcategories are presented in Table 7.

Table 7. *Frequencies related to the current status of content validity category and subcategories*

Current status of content validity	Face-to-face	Distance
	Frequency	Frequency
Evidence of content validity provided / high content validity	15	11
Low content validity	1	4

When Table 7 is analyzed, it is seen that academicians stated content validity was generally high in both face-to-face and distance exams. However, more academics stated that content validity was lower in distance exams than in face-to-face exams. Fifteen academics stated that content validity was high in face-to-face exams. For example, A3 stated that content validity was high in face-to-face exams, saying, "I can conduct exams with high content validity", and A10 said, "I can ask many questions. There is no problem in terms of content validity". Eleven academicians state that content validity is high in distance exams. The opinions of some of these academicians are as follows: (A4): "My content validity is high", (A5): "The content validity of the measurement and evaluation process is extremely high", (A13): "In both face-to-face education and distance education, I tried to cover all the subjects in the final exam by reducing the weight of the subjects covered in the midterm". On the other hand, while only one academician stated that content validity was low in face-to-face exams, four academicians stated it was low in distance exams. For face-to-face exams, academician A7 stated that "it is relatively low due to the course duration". For distance exams, A9 stated that "content validity may be lower compared to face-to-face exams. It was observed that academicians who expressed their opinions on the current status of content validity also mentioned the reasons for the current situation. The frequencies related to the category of reasons for the current status of content validity are given in Table 8.

Table 8. *Frequencies related to the category of reasons for the current status of content validity and its subcategories*

Reasons for the current status of content validity	Face-to-face	Distance
	Frequency	Frequency
<b>Evidence of content validity provided / high</b>		
Using a specification table	4	3
Sampling the topics well	4	5
Testing critical outcomes	1	0
Asking lots of questions and making exams	1	1
<b>Low content validity</b>		
Course duration	1	1
Using an item pool	0	1
Limitation of elements in the questions	0	1

When Table 8 is analyzed, it can be seen that the most frequently mentioned reasons for providing evidence for content validity in face-to-face exams are using a table of specifications ( $f=4$ ) and sampling the topics well ( $f=4$ ). For example, A11 stated, "First, I define the purpose of the test I am preparing and use a specification table to measure the intended content...". One of the academicians (A15) mentioned his practice as follows: "I prepare the specification table based on all learning outcomes, and all tools serve this purpose". Among the academicians who stated that they sampled the subjects well, A6 stated, "I prepare exam questions by weighting the topics I cover in the course according to the number of weeks the relevant topic is covered, the level of content covered, the subject titles and the duration of the topics covered although there is no specification table," and A14 stated, "In multiple-choice tests, I ask questions from each subject I show and try to distribute the questions equally". An academic (A7) mentioned the reasons for low content validity in face-to-face exams and



stated that content validity was relatively low due to the duration of the course. In addition, among the academicians who stated that the content validity was high or that they collected evidence of content validity, A9 mentioned that they tested critical outcomes, while A10 mentioned that they asked a large number of questions and conducted more than one exam. Similar to face-to-face exams, the most common reason for high content validity in distance exams was sampling the topics well ( $f=5$ ). This was followed by using the table of specifications ( $f=3$ ). Accordingly, it can be said that the use of the specification table decreased in the development of distance exams. Among the academicians who stated that they sampled the subjects well, A4 stated: "I try to sample all the subjects I have covered in a good way in the exam", while A14 stated that "I pay attention to the distribution of the questions equally in the homework according to the subjects I teach". In addition, A10 stated that in distance exams, as in face-to-face exams, he asked many questions and conducted more than one exam. Three different reasons were given for the low content validity in distance exams. These are limited course time, different exams for each student since the item pool is used, and the inability to use elements such as visuals, tables, and figures in distance exams.

#### iv. Exam Security

When the opinions of the academics were analyzed, it was seen that two categories related to the theme of exam security were formed. The categories related to the theme of exam security were determined based on the current security status and the precautions taken / that can be taken for security. The frequencies of the subcategories related to the current security status category are given in Table 9.

Table 9. *Frequencies for the current status category and subcategories*

Current exam security status	Face-to-face	Distance
	Frequency	Frequency
Classroom environment	2	0
No responsibility of the academician	4	2
High exam security	1	0
Low exam security	0	5
System security	0	1

When Table 9 is examined, most of the academicians ( $f=4$ ) mention that exam security is not their responsibility but that of the examiners. A6, one of the academicians who stated this view said, "Unfortunately, this issue is a bit out of my control as the management is responsible for determining the assistants in charge of the exam. Although I visit the classrooms at certain time intervals, I think this task is the responsibility of the assistant in charge of the exam". A12: "There is a proctor in each exam hall, and the exam supervisor is in the process. In other words, all possible conditions are tried to be provided for exam security". In the category of current exam security status, proctor responsibility is followed by the subcategory of classroom environment ( $f=2$ ). Regarding this subcategory, academician A2 stated, "Since our classrooms have very convenient environments, cheating attempts are almost zero." Finally, an academician stated that exam security is high in face-to-face exams. When the current

status of exam security in distance exams is examined, it is seen that academicians (f=5) mainly state that exam security is low. In this regard, A16: "There are many security gaps" and A17: "In distance education, we conduct the exam only for the sake of doing it. Some students really study for the exam, but unfortunately, there are also students who complete the exam by cheating... However, I can still understand whether students cheat or not from the exam results". The subcategory of low exam security for distance exams is followed by "no responsibility of the academician", which is a common subcategory with face-to-face exams (f=2). For example, A3 said, "I have very little chance to take security precautions individually". Finally, an academician states that the exam is as secure as the system's security in distance exams (A4). In addition to the current status of exam security, academicians also mentioned the measures taken and could be taken for exam security. The frequencies related to the category of measures taken / could be taken for security are presented in Table 10.

Table 10. *Frequencies related to the category and subcategories of precautions taken/could be taken for exam security*

Precautions taken/could be taken for exam security	Face-to-face	Distance
	Frequency	Frequency
<b>Ensuring security by the academician</b>		
Preparing different forms	7	0
Attending the exam in person	2	0
Psychological management	2	1
Utilizing in-class assessments	1	1
Collecting materials that provide advantages	1	0
Setting time limits	0	2
Avoid using questions that require memorization	0	1
Asking students to write the assignment by hand	0	1
<b>Taking systemic and management precautions</b>		
Mixing questions/options	0	6
Creating an item pool in the system	0	2
Enabling camera/audio activation	0	2
Increasing the number of proctors	2	0
Ensuring that students see the questions only once	0	1

In Table 10, when the opinions on the precautions taken or that could be taken for exam security in face-to-face exams are examined, it is seen that academicians (f=7) mostly state that they prepare different forms. For example, A1 stated that he diversified the test forms and changed the location of both questions and options by saying, "I prepare at least two forms, I change the location of both questions and options in the forms". A7 suggested "creating A and B forms for exams", and A10 mentioned the precaution he took as "I create A and B forms because there are crowded groups". In ensuring security by the academician, creating different forms in face-to-face exams is followed by attending the exam in person (the presence of the academicians themselves in the exam, f=2) and psychological management (f=2). In psychological management, academicians reported that they emphasized honesty and motivated them not to cheat. In addition, using in-class assessments predominantly and collecting materials such as written materials, cell phones, and calculators that will give students an advantage in answering questions are among the precautions that academicians have

taken or can take. Finally, increasing the number of proctors ( $f=2$ ) is among the systemic and administrative precautions for face-to-face exams. Academicians expect faculty administrations to appoint proctors and increase their number.

When the findings related to distance exams in the precautions taken / could be taken for exam security were analyzed, it was seen that academicians mostly expressed mixing the questions/options ( $f = 6$ ). For example, A8: "To prevent cheating, I design an exam where questions appear randomly and have a time limit"; A17: "As a precaution, we mix the questions and the answers". In these precautions that can be taken systemically and administratively, in addition to mixing the questions/options, they also mentioned withdrawing more items than the number of questions to be asked from the item pool in the system and turning students' cameras and voices on. When the precautions that academicians can take in distance exams are examined, they stated that they mostly ( $f=2$ ) set time limits on the questions and the exam. As a result, while academicians expressed more opinions about ensuring exam security in face-to-face exams, they mentioned more about the precautions that can be taken systematically and administratively in distance exams.

#### **v. Type of Application**

In this sub-problem, the opinions of academicians regarding their preferences for face-to-face and distance administration of measurement and evaluation processes were examined. It was found that 16 (94.11%) of the 17 academicians participating in the study preferred face-to-face exams. Only one academician (A1) stated that he preferred distance exams and mentioned the reasons for this preference as follows: "I prefer distance exams. Reading the exam, layout, and photocopying are no longer a workload. In addition, paper is saved, and there is no problem storing the exam documents." A1 generally emphasized the usefulness of distance exams and ecological sensitivity. Some academicians who prefer face-to-face exams also mentioned the reasons for their preference. For example, A4: "Because they are more reliable and valid", A6: "I definitely prefer face-to-face exams. I do not care about distance exams because the possibility of cheating is very high", A9: "I prefer face-to-face exams in terms of content validity and test security." A10:

I prefer face-to-face exams. At the undergraduate level, students still focus on passing the course. For this reason, they are trying to find a flaw in the exam software/process in online systems and cheat. These negatively affect the reliability and validity of the results obtained from the exam. Ultimately, all students, successful or unsuccessful, pass the exam. When exams are online, interest in the course decreases. We also have subjects that require computation in measurement. Paper and pencil exams are more practical for solving such computational questions.

A14 stated the reasons for preferring face-to-face exams: "Definitely face-to-face exams because students take it more seriously. Also, when they see that they do not know the subject and that subject is asked in KPSS, they are very motivated in the classroom. This situation motivates us as well". When

the reasons given by academicians are analyzed, it is seen that face-to-face exams are generally preferred in terms of reliability, validity, exam security, usefulness, and motivation.

Finally, A4, who stated that he preferred face-to-face exams, made additional explanations:

In fact, distance education has its advantages. My personal opinion is that they should be used together in the process. The distance education system is beneficial, especially for those who use alternative assessment approaches, or homework, etc. Distance education can be used in this period. However, if there is to be an exam, it must be face-to-face with a proctor, and a paper and pencil exam should be preferred.

In these explanations, A4 stated that although he preferred face-to-face exams, it is more useful to conduct complementary assessments distantly.

### **Conclusion and Discussion**

The study examined the opinions of academicians working in the field of measurement and evaluation, who carry out measurement and evaluation processes face-to-face and distantly, on both practices. It was determined that multiple-choice tests were much preferred in both application processes. Among the reasons for using multiple-choice tests, first, usefulness was stated, and second, high content validity and reliability were stated. It was determined that the usefulness of multiple-choice tests, especially in crowded groups, was effective in both applications. The high content validity since a large number of questions can be asked with multiple-choice items, the high reliability since it can be scored objectively, and the applicability to large groups in a short time are the features of this item type (Özçelik, 2010; Turgut, 1992). In addition, teachers at almost all educational levels are more familiar with traditional assessment tools, and one of the most commonly used traditional assessment tools is multiple-choice items (Anıl and Acar, 2008; Çakan, 2004, Gelbal and Kelecioğlu, 2007). Multiple-choice tests can also be easily implemented in the learning management system used in distance measurement and evaluation processes (Taşkıran, 2021). Secondly, academicians preferred mixed tests in face-to-face applications, while they preferred homework in distance processes. Among the reasons for the preference for mixed tests in face-to-face processes, it was stated that it provides the opportunity to ask questions at different levels. Since measurement tools have various advantages, mixed tests with varying measurement tools can be preferred. Short answers and written examinations were least preferred in face-to-face applications. The reason for the preference for written examinations was the ease of preparation in cases where multiple-choice exams could not be prepared. In distance education processes, the second most preferred assessment tool was homework, while written examinations and mixed tests were the least preferred. Academicians preferred distance measurement and evaluation assignments to provide student motivation and exam security. Students think that they learn new information with homework, and they can be motivated by thinking that they can improve their grades (Featherstone, 1985; cited in Kan, 2009). Similar to the results of this study, Karataş (2014) asked distance education students to rank the 11 most preferred assessment tools, and while homework was among

the top 5 tools, written examinations were ranked last. Usta et al. (2022) also stated in their study that experts' opinions were in favor of assigning homework in distance assessment processes. The study also showed that exam security was one of the reasons for preferring different assessment tools over face-to-face in-distance processes. Especially exam security is among the important risk issues in remote assessment and evaluation processes (Acar Güvendir and Özer Özkan, 2021; Atılğan et al., 2020; Bozkurt and Uçar, 2018; James, 2016; Shraim, 2018). The literature has proposed different applications to ensure exam security (Al-Shalout, Rasmi, and Hassan, 2021; Bedford, Gregg, and Clinton, 2011; Bozkurt and Uçar, 2018; James, 2016). Authentication classifications can be used. Invigilator or electronic invigilator software combined with biometrics applications can be used. Fingerprint readers and 360-degree cameras can be used. The student may be asked one or more special security questions during the exam. In addition, studies to determine ethical rules can also be carried out (Bozkurt, 2020).

Within the scope of the research, reliability and validity, which are essential issues of measurement and evaluation, were also examined. Most academicians stated that face-to-face measurement and evaluation processes were highly reliable. On the contrary, they indicated that measurement and evaluation processes had low reliability in distance processes. The number of questions, group heterogeneity, ensuring exam security, and using an item pool with item statistics known were among the reasons for high reliability in face-to-face education. In distance education, on the other hand, it was stated that reliability was low, primarily due to concerns about exam security. One of the striking issues in the reliability theme was that the use of technology negatively affected reliability in distance processes. Similar to the results of this study, Sharim (2018) reported that 70.5% of the participants found the reliability of the test low due to the technology used. Although there is a positive reflection of the use of technology in stages such as preparation, application, and correction in the exam process (Hack and Tarouco, 2000; cited in Karadağ, 2021), it also creates some concerns. Technological infrastructure, security issues, and competence to use technology may be among these concerns. In the study by James (2016) with university students, students stated that the fear of having technology problems was the reason why they did not prefer online exams. Therefore, the concern of having problems with technology may also negatively affect students' motivation.

Increasing the number of questions in both applications was stated as one of the precautions that could be taken to increase the reliability of the test. Increasing the number of questions will positively affect the reliability as it will increase the sensitivity of the measurement tool. In the study of Usta et al. (2022), it was stated that experts offered suggestions for increasing the number of questions related to the theme of measurement and evaluation in online exams and providing diversity in measurement tools. Within the scope of the research, it has been stated that unlike face-to-face and distance processes, the ability of the students to ask the subject they do not understand may positively affect reliability. In distance measurement and evaluation processes, it was stated that precautions such as ensuring exam security, using different measurement tools, avoiding questions that require

memorization, and creating an item pool would contribute positively to reliability. Chaudhary and Dey (2013), in their study examining the assessment process in open and distance education, stated that evaluation should be carried out through a variety of tools. They suggested tools such as written exams, short answer and objective tests, oral exams, practical exams, field-based practice and evaluation, and project and performance evaluation. In their study, Yağcı, Ekiz, and Gelbal (2015) stated that using multiple-choice items with more than one correct answer could also increase reliability. When the study findings are evaluated in general, it can be said that measurement and evaluation processes carried out distantly are more prone to measurement errors.

In the analysis of content validity in face-to-face and distance processes, evidence was provided that content validity was achieved in both face-to-face and distance measurement and evaluation processes, and it was stated to be high. However, the number of those who stated that content validity was low in distance assessment and evaluation processes was higher than in face-to-face assessment and evaluation processes. Sampling the topics by item is among the reasons for high content validity in both applications. In the research, it was also stated that the limitations in preparing exams in distance processes, such as not being able to upload elements such as graphics, figures, and tables in the content of the items to the system, caused lower content validity. This is considered a limitation arising from the system used in distance education. However, when evaluated in general, it can be concluded that content validity can be achieved in both applications.

Another theme addressed within the scope of the research was security, and the opinions of academicians were discussed for both applications. Most academicians stated they had no responsibility for face-to-face exams because the proctors were on duty. In distance measurement and evaluation processes, it was mostly stated that exam security was low. They also stated precautions academicians can take to provide exam security in face-to-face applications.

In their study, Al-Shalout et al. (2021) examined techniques to ensure the security of distance exams. They noted that despite all the precautions that can be taken to increase security, electronic applications cannot compete with exams in school or university halls. Another noteworthy finding in the security theme is the preparation of different forms. Especially in preventing cheating behavior, creating different forms in crowded classes may be preferable. Content balancing should be ensured when preparing different forms. However, in cases where item difficulty and item discrimination cannot be prepared equally, forms prepared based on changing the location of the questions can also be used. Mostly, using different forms in which the order of the questions is changed is preferred. In their study, Yağcı et al. (2015) also stated that asking different questions to each student or presenting the same questions to students with different correct answers can be used to increase test security. In the face-to-face application, the majority stated the necessity of systemic and administrative precautions. Whitley (1998), in the research on cheating behavior of high school students, suggested that absolute

assessment should be preferred in assessments; in small classes, students should be seated in a large area, and if this is not possible, different test forms should be prepared, and students should be carefully monitored throughout the exam as a precaution to cheating behavior.

Finally, academicians were asked about their preferences for face-to-face and distance measurement and evaluation processes. All academicians except one stated that they preferred face-to-face applications. They noted that face-to-face processes were chosen to ensure the measurement tools' validity, reliability, and usefulness. In general, it is seen that both measurement and evaluation processes have advantages and limitations compared to each other. However, it was also emphasized that face-to-face measurement and evaluation processes are preferable regarding the psychometric properties of measurement tools. Based on the study's results, it can be recommended that researchers compare the results of both practices by conducting experimental research. Studies can also be carried out to compare the processes by addressing themes different from those addressed within this research scope.

## Kaynaklar

- Acar Güvendir, M., & Özer Özkan, Y. (2021). Uzaktan eğitimin değerlendirmeye yansımaları: çevrim içi sınavlar mı sınıf içi sınavlar mı?. *Journal of Digital Measurement and Evaluation Research*, 1(1), 22-34. <https://doi.org/10.29329/dmer.2021.285.2>
- Al-Shalout, M. I., Rasmi, M., & Hassan, M. A. (2021). Evaluation of E-exam during Covid-19. *Psychology and Education Journal*, 58(1), 4604-4612.
- Anıl, D. ve Acar, M. (2008). Sınıf öğretmenlerinin ölçme değerlendirme sürecinde karşılaştıkları sorunlara ilişkin görüşleri. *Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi*, 5(2), 44-62.
- Arslan, E. (2022). Nitel araştırmalarda geçerlilik ve güvenilirlik. *Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 51, 395-407. <https://doi.org/10.30794/pausbed.1116878>
- Atılğan, B., Tarı, O. E., Özdemir, B. N., Aktar, İ., Güneş, M., ..... Sayek, İ. (2020). Tıp fakültesi öğrencilerinin gözünden acil uzaktan öğretim sürecinin değerlendirilmesi. *Sürekli Tıp Eğitimi Dergisi*, 29(6), 396-406. <https://doi.org/10.17942/sted.837551>
- Baran, H. (2020). Açık ve uzaktan eğitimde ölçme ve değerlendirme. *Açıköğretim Uygulamaları ve Araştırmaları Dergisi*, 6(1), 28-40.
- Başol, G. (2018). *Eğitimde ölçme ve değerlendirme* (5. Baskı). Pegem.
- Bedford, D. W., Gregg, J. R., & Clinton, M. S. (2011). Preventing online cheating with technology: a pilot study of remote proctor and an update of its use. *Journal of Higher Education Theory and Practice*, 11(2), 41-58.
- Bıkmaz Bilgen, Ö. (2019). Tamamlayıcı ölçme ve değerlendirme teknikleri I: Performans değerlendirme. İçinde N. Doğan (Ed.) *Eğitimde ölçme ve değerlendirme* (s. 182-214). Pegem.
- Bozkurt, A. (2020). Koronavirüs (Covid-19) pandemi süreci ve pandemi sonrası dünyada eğitime yönelik değerlendirmeler: *Yeni normal ve yeni eğitim paradigması*. *Açıköğretim Uygulamaları ve Araştırmaları Dergisi*, 6(3), 112-142. <https://dergipark.org.tr/en/download/article-file/1215818> sayfasından erişilmiştir.
- Bozkurt, A., & Uçar, H. (2018). E-Öğrenme ve e-sınavlar: Çevrimiçi ölçme değerlendirme süreçlerinde kimlik doğrulama yöntemlerine ilişkin öğrenen görüşlerinin incelenmesi. *Mersin Üniversitesi Eğitim Fakültesi Dergisi*, 14(2), 745-755. <https://doi.org/10.17860/mersinefd.357339>
- Çakan, M. (2004). Öğretmenlerin ölçme-değerlendirme uygulamaları ve yeterlik düzeyleri: İlk ve ortaöğretim. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 37(2), 99-114.
- Chaudhary, S., & Dey, N. (2013). Assessment in open and distance learning system (ODL): A Challenge. *Open Praxis*, 5(3), 207-216.
- Doğan, N. (2009). Ölçme araçlarını sınıflama çabaları: İçinde H. Atılğan, A. Kan ve N. Doğan (Ed.) *Eğitimde ölçme ve değerlendirme* (s. 119-144). Anı.



- Doğan, N. (2019). İnsan davranışları ve ölçme araçlarını sınıflama. İçinde N. Doğan (Ed.) *Eğitimde ölçme ve değerlendirme* (s. 76-111). Pegem.
- Eroğlu, F., & Kalaycı, N. (2020). Üniversitelerdeki zorunlu ortak derslerden yabancı dil dersinin uzaktan eğitim uygulamasının değerlendirilmesi. *Türk Eğitim Bilimleri Dergisi*, 18(1), 236-265. <https://doi.org/10.37217/tebd.683250>
- Gelbal, S., & Kelecioğlu, H. (2007). Öğretmenlerin ölçme ve değerlendirme yöntemleri hakkındaki yeterlik algıları ve karşılaştıkları sorunlar. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 33, 135-145.
- Gültekin, S. (2017). Performans dayanaklı değerlendirme. İçinde N. Çıkrıkçı (Ed.) *Eğitimde ölçme ve değerlendirme* (s. 223-256). Anı.
- James, R. (2016). Tertiary student attitudes to invigilated, online summative examinations. *International Journal of Educational Technology in Higher Education*, 13(1), 1-13.
- Kan, A. (2009). Ödev ve projeler. İçinde H. Atılgan, A. Kan ve N. Doğan (Ed.) *Eğitimde ölçme ve değerlendirme* (s: 269-292). Anı.
- Karadağ, N. (2014). *Açık ve uzaktan eğitimde ölçme ve değerlendirme: Mega üniversitelerdeki uygulamalar* (Yayımlanmamış Doktora Tezi). Sosyal Bilimler Enstitüsü, Anadolu Üniversitesi, Eskişehir
- Karadağ, N. (2021). Açıköğretim Sistemi'nde çevrimiçi sınav uygulamasının sınav hazırlama sürecine etkisi. *Açıköğretim Uygulamaları ve Araştırmaları Dergisi*, 7(1), 45-60. <https://dergipark.org.tr/tr/download/article-file/1498588> sayfasından erişilmiştir.
- Karataş, S. (2003). Yüz yüze ve uzaktan eğitimde öğrenme deneyimlerinin eşitliği. *Eğitim Bilimleri ve Uygulama*, 2(3), 91-104. [http://www.ebuline.com/pdfs/3Sayi/ebu3\\_7.pdf](http://www.ebuline.com/pdfs/3Sayi/ebu3_7.pdf) sayfasından erişilmiştir.
- Kilmen, S. (2017). Ölçme ve değerlendirmede temel kavramlar. İçinde N. Çıkrıkçı (Ed.) *Eğitimde ölçme ve değerlendirme* (s. 25-56). Anı.
- Kutlu, Ö., Doğan, C. D., & Karakaya, İ. (2010). *Ölçme ve değerlendirme performans ve portfolyoya dayalı durum belirleme*. Pegem.
- Meriam, S. B. (2009). *Qualitative research a guide to design and implementation*. Jossey-Bass.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Sage.
- Ocak, G., & Karakuş, G. (2022). Üniversite öğrencilerinin çevrimiçi sınavlara yönelik tutumlarına ilişkin ölçek geliştirme çalışması. *Afyon Kocatepe Üniversitesi Sosyal Bilimler Dergisi*, 24(1), 66-86. <https://doi.org/10.32709/akusosbil.887141>
- Özçelik, D. A. (2010). *Okullarda ölçme ve değerlendirme: Öğretmen el kitabı*. Pegem.
- Popham, J. (1999). *Classroom assessment: What teachers need to know?* (2. ed.). Allyn & Bacon.

- Sarı, H. (2020). Evde kal döneminde uzaktan eğitim: Ölçme ve değerlendirmeyi neden karantinaya almamalıyız?. *International Journal of Scholars in Education*, 3(1), 121-128. <https://dergipark.org.tr/tr/download/article-file/1181932> sayfasından erişilmiştir.
- Shraim, K. (2019). Online examination practices in higher education institutions: learners' perspectives. *Turkish Online Journal of Distance Education*, 20(4), 185-196. <https://doi.org/10.17718/tojde.640588>.
- Simonson, M., Zvacek, S. M., & Smaldino, S. (2012). *Teaching and learning at a distance: Foundations of distance education* (3rd ed.). Prentice Hall.
- Şahin, M. G. (2019). Performansa dayalı değerlendirme. İçinde B. Çetin (Ed.), *Eğitimde ölçme ve değerlendirme* (s. 219-268). Pegem.
- Taşkıran, A. (2021). Açık ve uzaktan öğrenme süreçlerinin yönetim, öğrenme, teknoloji ve değerlendirme boyutları. *Açıköğretim Uygulamaları ve Araştırmaları Dergisi*, 7(4), 80-98. <https://doi.org/10.51948/auad.984969>
- Turgut, F. (1992). *Eğitimde ölçme ve değerlendirme metotları*. ÖSYM.
- Usta, İ., Kılınç, H., & Okur, M. R. (2022). Açık ve uzaktan öğrenme ortamlarında uygulanan ölçme ve değerlendirme sürecine ilişkin alan uzmanlarının görüşleri: bir durum çalışması. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, 42(1), 167-200. <https://doi.org/10.17152/gefad.944272>
- Whitley, B. E. (1998). Factors associated with cheating among college students: A review. *Research in Higher Education*, 39(3), 235-274. <https://doi.org/10.1023/A:1018724900565>
- Yağcı, M., Ekiz, H., & Gelbal, S. (2015). Yeni bir çevrimiçi sınav modeli geliştirilmesi ve uygulanması. *Ahi Evran Üniversitesi Kırşehir Eğitim Fakültesi Dergisi*, 16(1), 269-288. <https://dergipark.org.tr/tr/download/article-file/1489321> sayfasından erişilmiştir.
- Yıldırım, A., & Şimsek, H. (2011). *Sosyal bilimlerde nitel araştırma yöntemleri*. Anı.