



## Bolu Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi (BAİBÜEFD)


Bolu Abant İzzet Baysal University  
Journal of Faculty of Education



2023, 23(1), 249–271. <https://dx.doi.org/10.17240/aibuefd.2023.-1192209>

### E-Değerlendirmeye Yönelik Tutum Ölçeğinin Madde Yanlılığının İncelenmesi\*

An Investigation of Item Bias for the Attitude Scale Towards E-Assessment

Ceylan Gündeğer<sup>1</sup> 

Geliş Tarihi (Received): 20.10.2023

Kabul Tarihi (Accepted): 20.01.2023

Yayın Tarihi (Published): 25.03.2023

**Öz:** Covid-19 ile tekrar gündeme gelen elektronik değerlendirme (e-değerlendirme) kavramının, çağın bir gereksinimi olarak ilerleyen yıllarda epey yaygınlaşacağı düşünülmektedir. Bu açıdan, e-değerlendirme kavramına ilişkin tutumların incelenmesinin günümüz ve sonrasında önem arz edeceği söylenebilir. Türkiye’de kullanılan e-değerlendirmeye yönelik tutum ölçekleri dikkate alındığında, ölçeklerin geçerliğine ve güvenilirliğine ilişkin kanıtların ortaya koyulduğu görülmektedir. Ancak bu araştırmalarda gruplar arası karşılaştırmalar yapılmış olsa da, ölçme değişmezliğinin ne düzeyde sağlandığı, değişen madde fonksiyonu (DMF) gösteren ve yanlılığa işaret ederek geçerliği düşüren herhangi bir maddenin bulunup bulunmadığı incelenmemiştir. Geçerliğe ilişkin daha fazla kanıt sunması bakımından önem arz eden bu karma yöntem araştırmasının amacı, Bahar tarafından geliştirilmiş, Bahar ve Asil tarafından revize edilmiş, e-Değerlendirmeye Yönelik Tutum Ölçeği’nin, cinsiyet ve kişisel bir bilgisayara sahip olma değişkenlerine göre, ölçme değişmezliği, DMF ve madde yanlılığı açısından incelenmesidir. Araştırmanın nicel kısmı 537 lisans öğrencisinden, nitel kısmı ise beş alan uzmanından oluşmaktadır. Araştırmanın nicel kısmının birinci basamağında, elde edilen veri, analiz varsayımları kontrol edilerek, bağımsız değişkenler temelinde ölçme değişmezliği bakımından sınanmıştır. Her iki bağımsız değişkende de metrik değişmezliğin sağlanamıyor oluşu sebebiyle, ikinci aşamada veri setine POLY-SIBTEST ile MantelTest uygulanmış ve DMF içeren maddeler belirlenmiştir. Son olarak nitel kısımda, iki yöntem sonucu yüksek düzeyde DMF gösterdiği belirlenen maddelere ilişkin uzman görüşlerine başvurularak maddelerin yanlılık gösterip göstermediği tartışılmış, bu bağlamda araştırmacılara ve uygulayıcılara öneriler sunulmuştur.

**Anahtar Kelimeler:** Ölçme değişmezliği, değişen madde fonksiyonu, yanlılık, e-değerlendirme, tutum.

&

**Abstract:** It is thought that the concept of electronic assessment (e-assessment), which has become a current issue with Covid-19, will gain widespread use in the upcoming years as a requirement of the modern age. In this respect, it can be suggested that analyzing attitudes toward e-assessment will be important in the present and the future. Considering the attitude scales for e-assessment used in Turkey, it can be observed that some relevant evidence has been demonstrated for the validity and reliability of the scales. However, although inter-group comparisons were performed in these studies, they did not investigate at what level measurement invariance is achieved; or whether any items were demonstrating differential item functioning (DIF), and reducing validity by indicating bias. This mixed-method study is significant in terms of providing further evidence of validity and it aims to examine the e-Assessment Attitude Scale, developed by Bahar and revised by Bahar & Asil, in terms of measurement invariance and differential item functioning according to the variables of gender and possession of a personal computer. The quantitative part of the study consisted of 537 undergraduate students and the qualitative part consisted of eight field experts. During the first step of the quantitative part of the study, the data obtained were tested for measurement invariance based on independent variables by checking the analysis assumptions. Due to the fact that metric invariance could not be achieved in both independent variables, POLY-SIBTEST and MantelTest were applied to the data set in the second step, and items demonstrating DIF were determined. Finally, in the qualitative part, the researchers discussed whether the items were biased or not by referring to the expert opinions on the items that demonstrate a high level of DIF according to the results of both methods, and relevant recommendations were presented to the researchers and practitioners.

**Keywords:** Measurement invariance, differential item functioning, bias, e-assessment, attitude.

**Atıf/Cite as:** Gündeğer, C. (2023). E-Değerlendirmeye yönelik tutum ölçeğinin madde yanlılığının incelenmesi. *Bolu Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 23(1), 249-271. <https://dx.doi.org/10.17240/aibuefd.2023.-1192209>

**İntihal-Plagiarism/Etik-Ethic:** Bu makale, en az iki hakem tarafından incelenmiş ve intihal içermediği, araştırma ve yayın etiğine uyulduğu teyit edilmiştir. / This article has been reviewed by at least two referees and it has been confirmed that it is plagiarism-free and complies with research and publication ethics. <https://dergipark.org.tr/tr/pub/ijaws>

**Copyright** © Published by Bolu Abant İzzet Baysal University– Bolu

\* Bu araştırmanın bir kısmı, 1-3 Nisan 2022 tarihlerinde Ankara’da düzenlenen VI. Ankara Uluslararası Bilimsel Araştırmalar Kongresi’nde sözlü bildiri olarak sunulmuştur.

<sup>1</sup> Sorumlu Yazar: Dr. Öğr. Üyesi Ceylan Gündeğer, Aksaray Üniversitesi, Eğitim Bilimleri Bölümü, [cgundege@gmail.com](mailto:cgundege@gmail.com), ORCID: <https://orcid.org/0000-0003-3572-1708>

## 1. GİRİŞ

Covid-19 sebebiyle eğitim süreci ülkemizde ve tüm dünyada çevrim içi ortamda gerçekleşmiştir. Eğitimde ölçme ve değerlendirme uygulamaları da benzer şekilde bu ortama taşınmış ve elektronik değerlendirme (e-değerlendirme) kavramı yine bu zaman diliminde önem kazanmıştır. Türkiye’de e-YDS ile Bilgisayar Tabanlı Test (BBT) uygulamaları Pandemi öncesinde de gündeme gelen bir konu olmuştur. Özellikle, kısa bir geçmişe sahip olmasına rağmen Otomatik Madde Üretimi (OMÜ), test geliştirme ve geçerlik için heyecan verici ve yenilikçi teknolojilerin kullanılmasını vaat etmektedir (Gierl & Haladyna, 2013). BBT ile OMÜ’nin avantajları ve çağın gereksinimleri dikkate alındığında, ilerleyen yıllarda e-değerlendirme uygulamalarının daha da yaygınlaşacağı öngörülebilmektedir. Bu açıdan bakıldığında, öğrencilerin ve hatta geleceğin öğretmenlerinin e-değerlendirmeye yönelik tutumlarının farklı değişkenler temelinde incelenmesinin günümüz ve sonrasında önem arz etmeye devam edecek konular arasında olduğu söylenebilir.

Tutum, algı, ilgi gibi bireylere özgü psikolojik özellikler hakkında bilgiler ölçekler yardımıyla dolaylı yoldan edinilir. Ölçeklerin geçerli ve güvenilirliği yüksek ölçümler sağlaması beklenir. Güvenirlik ölçüm sonuçlarının tesadüfi hatalardan arınlık düzeyini gösterirken; geçerlik, daha geniş bir kavram olmakla birlikte, ölçme aracının amacına uygunluğuna ilişkin kanıt toplama sürecini ifade etmektedir (Messick, 1995). Ölçme aracının geçerliğine ilişkin kanıt toplama aşamasında, ölçme aracının değişmezlik ve yanlılık bakımından incelenmesi bir gereklilik olarak düşünülebilir. Cheung ve Rensvold (2002), gruplar karşılaştırılacağı zaman, ölçme değişmezliğinin kritik bir öneme sahip olduğunu belirtmektedir. Eğer ölçme değişmezliği sağlanamazsa, gruplar arasında ortaya çıkan farklılıkların gerçek farklılıklardan mı yoksa ölçek maddelerine verilen farklı psikometrik cevaplardan mı kaynaklandığı bilinemeyecek ve sonuçlar yorumlanamayacaktır (Cheung & Rensvold, 2002).

İki gruba ait gözlenen ölçek puanlarının karşılaştırılmasında, gruplarda yer alan bireylerin ölçeği aynı şekilde anlaması gerekmektedir. Horn ve McArdle (1992) gruplar arası geçerli karşılaştırmaların yapılabilmesi için ölçümlerin psikometrik eş değerliğinin (değişmezliğinin) ortaya koyulmasının gerekli olduğunu ifade etmişlerdir (akt: Meade & Lautenschlager, 2004). Örneğin, bir tutum ölçeğinden veya başarı testinden elde edilen öğrenci puanları üzerinden cinsiyet değişkeni dikkate alınarak bir karşılaştırma yapılacaksa, öncelikle ölçeğin kız ve erkek öğrenci gruplarında ölçme değişmezliğini sağladığına ilişkin kanıtların sunulması gerekmektedir. Bir başka ifadeyle, ölçme değişmezliğinin gruplar arasında sağlandığının ortaya koyulması, gruplar arası karşılaştırmaların yapılabileceğine ve bu karşılaştırmaların anlamlı olabileceğine işaret edecektir. Ölçme değişmezliğini test etmede doğrulayıcı faktör analitik yöntemler işe koşulmaktadır (Meade & Lautenschlager, 2004). Meredith’e (1993) göre, ölçme değişmezliği bir hiyerarşi içerisinde dört aşamada gerçekleşmektedir. Bunlar Yapısal, Metrik, Skalar (Güçlü) ve Katı değişmezliktir. Bu aşamalardan birinin sağlanamaması durumunda diğer aşamadaki değişmezlik türü test edilmemektedir.

Yapısal değişmezlikte, her grupta aynı maddelerin ortak faktörlerle ilişkili olup olmadığı test edilmektedir. Yapısal değişmezliğin sağlanması, madde kümelerinin her grupta aynı şekilde yer aldığına işaret etmekte fakat gruplardan kestirilen parametrelerin değişmezliğine kanıt sunmamaktadır. Metrik değişmezlikte grupların ortak faktörleri benzer şekilde algılayıp algılamadığı; faktör yüklerinin gruplar arasında eşitliği sınırlanmaktadır. Metrik değişmezliğin sağlanamaması durumunda iki olasılıktan söz edilir. Bunlardan biri, ortak faktörlerden biri ya da birkaçı veya madde kümelerinden en az biri, gruplarda farklı bir anlama sahip olabilir. İkinci olasılık ise faktör yükü kestirimlerinin yanlı olmasıdır. Skalar değişmezlik aşamasında ise faktör yükleriyle birlikte regresyon sabitlerinin eşitliği de test edilmektedir. Skalar değişmezliğin sağlanması durumunda madde ortalamalarının gruplar arasında eşit olduğu yorumu yapılır ve son aşamada katı değişmezlik incelenir. Katı değişmezlikte ise, hata varyanslarının gruplarda eşit olup olmadığı incelenmektedir. Skalar değişmezliğin sağlanması; bir başka deyişle faktör yüklerinin ve regresyon sabitlerinin gruplarda değişmez olması, (1) kestirilen faktör ortalamalarındaki grup farklılıklarının yansız olacağına ve (2) gözlenen puan ortalamalarındaki farklılıkların da direkt olarak grup

farklılıklarıyla ilişkili olacağına işaret edecektir (Gregorich, 2006). Buna göre skalar değişmezliğin sağlanamaması durumunda madde yanlılığının gündeme geleceği yorumu yapılabilir.

Yanlılık, test ve madde yanlılığı olarak iki farklı şekilde ele alınmaktadır. Madde yanlılığı, farklı kültürden, cinsiyetten veya etnik kökenden gelen aynı yetenekteki bireylerin maddede farklı düzeyde performans göstermesi olarak tanımlanmaktadır (Hambleton & Rodgers, 1994). Dolayısıyla sistematik bir hatadır ve geçerliği düşürür. Madde yanlılığın incelenmesinde değişen madde fonksiyonu (DMF) analizleri işe koşulmaktadır. DMF, aynı yetenekteki farklı grupların (örneğin benzer tutum düzeyindeki kız ve erkek öğrencilerin) ölçek maddesine/ifadesine tepkide bulunma olasılığı farklılık gösterdiğinde ortaya çıkmaktadır. DMF'nin belirlenmesinde kullanılan yöntemler genel olarak, Klasik Test Kuramı (KTK) ve Madde Tepki Kuramı (MTK) temelinde ayrışsa da, ölçme aracındaki maddelerin ne şekilde puanlandığında göre de farklılık göstermektedir. İkili puanlamaya dayanan bir ölçekte DMF analizleri Mantel Haenszel, Lojistik Regresyon gibi yöntemlerle yapılırken; çoklu puanlanan maddelerin bulunduğu ölçeklerde DMF gösteren maddelerin belirlenebilmesi amacıyla POLY-SIBTEST, MantelTest, ordinal lojistik regresyon gibi yöntemlerden yararlanılmaktadır. Bu çalışmada, "Hiç katılmıyorum"dan (1) "Tamamen katılıyorum"a (5) kadar, 5'li Likert tipinde puanlanan bir tutum ölçeğinin incelenmesi sebebiyle çoklu puanlamaya dayalı yöntemlerden POLY-SIBTEST ile MantelTest yöntemleri işe koşulmuş olup yöntemlere veri analizi bölümünde yer verilmiştir. Eğer bir maddede DMF gözlemlenmezse bu maddenin yanlılık içermediği söylenmektedir. Ancak maddede DMF olduğu tespit edilirse maddenin yanlılık gösterme ihtimalinin olduğu düşünülür ve maddeye tamamlayıcı madde yanlılığı analizleri uygulanır (Zumbo, 1999). Bu aşamada uzman görüşlerine başvurulmakta; gruplar arasındaki farklılığın madde etkisinden mi yanlılıktan mı kaynaklandığı ortaya koyulmaya çalışılmaktadır.

E-değerlendirme kapsamında Türkiye'de geliştirilen ve kullanılan tutum ölçekleri, Gül vd. (2015) tarafından geliştirilmiş Online Değerlendirmeye Yönelik Tutum Ölçeği (ODYTÖ); Başaran vd. (2016)'ın geliştirdiği Web-Tabanlı Sınav Sistemine (Moodle) Yönelik Tutum Ölçeği (WTSSYT) ile Bahar (2014) tarafından geliştirilmiş, Bahar ve Asil (2018) tarafından revize edilmiş E-Değerlendirmeye Yönelik Tutum Ölçeği'dir (EDYTÖ). Gül vd. (2015) tarafından geliştirilmiş ODYTÖ'nin Online Değerlendirme Uygulamalarının Kullanımı isimli alt boyutuna ait güvenilirlik katsayısının 0.62 şeklinde hesaplanmış olması; Başaran vd. (2016) tarafından geliştirilen WTSSYT'nin ise doğrudan Moodle sistemine yönelik hazırlanmış olması bu çalışmada Bahar (2014) tarafından geliştirilmiş ve Bahar ve Asil (2018) tarafından revize edilmiş EDYTÖ'nün incelenmesine temel teşkil etmiştir. Bahsedilen çalışmalarda gruplar arası karşılaştırmalar yapılmış olsa da, ölçme değişmezliğinin ne düzeyde sağlandığına, DMF veya yanlılık gösteren madde olup olmadığına ilişkin herhangi bir incelemenin yer almadığı dikkat çekmektedir (Bahar, 2014; Bahar & Asil, 2018; Başaran vd., 2016; Gül vd., 2015).

### 1.1. Araştırmanın amacı

Bu çalışmada, Bahar (2014) tarafından geliştirilmiş ve Bahar ve Asil (2018) tarafından revize edilmiş, e-Değerlendirmeye Yönelik Tutum Ölçeği'nin, cinsiyet ve kişisel bir bilgisayara sahip olma (PC) değişkenlerine göre ölçme değişmezliğinin, değişen madde fonksiyonunun ve madde yanlılığının incelenmesi amaçlanmıştır. Bu amaç doğrultusunda aşağıdaki alt problemlere cevap aranmıştır:

- 1- e-Değerlendirmeye Yönelik Tutum Ölçeği cinsiyet ve PC'ye göre ölçme değişmezliğini ne düzeyde sağlamaktadır?
- 2- e-Değerlendirmeye Yönelik Tutum Ölçeği'nde yer alan maddelerde, cinsiyet ve PC'ye göre POLY-SIBTEST yöntemi işe koşulduğunda değişen madde fonksiyonu bulunmakta mıdır?
- 3- e-Değerlendirmeye Yönelik Tutum Ölçeği'nde yer alan maddelerde, cinsiyet ve PC'ye göre MantelTest yöntemi işe koşulduğunda değişen madde fonksiyonu bulunmakta mıdır?
- 4- İki yöntem sonucunda, yüksek düzeyde değişen madde fonksiyonu içerdiği tespit edilen maddelere ilişkin uzman görüşleri nasıldır?

## 1.2. Araştırmanın önemi

Alanyazın incelendiğinde, değişmezlik, DMF veya yanlılık temelinde, Türkiye’de genellikle büyük ölçekli sınavlar üzerinden kültürler arası eş değerliğin incelendiği veya simülasyon çalışmalarında farklı koşullar altında farklı yöntemlerin karşılaştırıldığı görülmektedir (Arıkan Akın, 2015; Arıkan vd., 2015; Asil & Gelbal, 2012; Atalay, 2010; Ayan, 2011; Bakan Kalaycıoğlu & Kelecioğlu, 2011; Başokçu & Öğretmen, 2013; Çepni, 2011; Demir, 2013; Doğan & Öğretmen, 2008; Erdem Keklik, 2012; Gök vd., 2010; Gür, 2019; Karakaya & Kutlu, 2012; Korkmaz, 2005; Öğretmen, 2006; Sırgancı, 2012; Tiryaki, 2019; Tunç, 2016; Ünsal Özberk & Koç, 2017; Yandı, 2017). Bu araştırmaların birçoğu ikili puanlanan maddeler üzerinden gerçekleştirilmiş olup çoklu puanlanan maddelere ilişkin daha az sayıda DMF araştırması olduğu söylenebilir (Arıkan Akın, 2015; Atalay, 2010; Bilican Demir, 2014; Gür, 2019; Korkmaz, 2005; Kuzu, 2021; Sırgancı, 2012; Tiryaki, 2019; Ünsal Özberk & Koç, 2017; Yandı, 2017). Çoklu puanlamaya dayanan maddelerin ise yalnızca çok azının gerçek veri seti üzerinden psikolojik bir özelliği ölçmeyi amaçlayan ölçek maddelerinin yanlılığını ele aldığı dikkati çekmektedir (Başokçu & Öğretmen, 2013; Ünsal Özberk & Koç, 2017; Yıldız, 2022). Bu açıdan bakıldığında araştırmanın, geçerliği etkileyen bir faktör olarak düşünülebilecek yanlılığa ilişkin kanıt sunması bakımından alana katkı sağlayacağı ve gerçek veri seti üzerinden gerçekleşen ampirik bir çalışma olması ile de önem arz ettiği söylenebilir.

## 2. YÖNTEM

### 2.1. Araştırmanın modeli

Bu araştırmada, e-Değerlendirmeye Yönelik Tutum Ölçeği’nin, cinsiyet ve kişisel bir bilgisayara sahip olma değişkenleri temelinde ölçme değişmezliğinin incelenmesi, ölçekte değişen madde fonksiyonu (DMF) gösteren maddelerin belirlenmesi; DMF içeren maddelerin uzman görüşleri doğrultusunda yanlılık bakımından ele alınması amaçlanmıştır. Bu araştırmada, nicel ve nitel yöntemler birlikte işe koşulduğundan araştırmanın karma türde olduğu söylenebilir. Karma yöntem araştırmalar, hem nicel hem nitel verinin toplandığı araştırmalardır (Creswell, 2014). Bu araştırmada, öncelikle değişmezlik ve DMF analizleri ile nicel alt problemlerin çözümüne, ardından yüksek düzeyde DMF içerdiği belirlenen maddelere yönelik uzman görüşleriyle nitel alt problemin incelenmesine yer verilmiştir. Bir başka ifadeyle araştırmada önce nicel aşama tamamlanmış, ardından nicel kısım temelinde nitel aşamaya geçilmiştir. Nicel ve nitel aşamaların sırayla uygulanmış olması ve iki sonuç seti bütünleştirilerek çıkarımlar yapılmış olmasına dayanılarak araştırmanın sıralı “Nic -> Nit” desende yürütüldüğü söylenebilir (Yalçın, 2018).

### 2.2. Araştırmanın çalışma grubu

Araştırmanın nicel boyutundaki çalışma grubunu, 2021-2022 eğitim öğretim yılında, bir devlet üniversitesinin eğitim fakültesinde lisans düzeyinde öğrenim görmekte olan 537 öğretmen adayı oluşturmaktadır. Çalışma grubuna ilişkin betimsel istatistikler Tablo 1’de sunulmuştur.

**Tablo 1.**

*Çalışma Grubunun Betimsel İstatistikleri (N=537)*

		f	%
<b>Cinsiyet</b>	Kadın	406	75.6
	Erkek	131	24.4
<b>PC</b>	Var	371	69.1
	Yok	166	30.9
<b>Bölüm</b>	Beden Eğitimi	32	6.0
	Fen Bilgisi Eğitimi	15	2.8
	İngilizce Öğretmenliği	10	1.9
	Matematik Eğitimi	49	9.1

**Tablo 1. Devamı**

*Çalışma Grubunun Betimsel İstatistikleri (N=537)*

		f	%
<b>Bölüm</b>	Müzik Öğretmenliği	9	1.7
	Okul Öncesi Öğretmenliği	90	16.8
	Rehberlik ve Psikolojik Danışmanlık	167	31.1
	Resim-İş Öğretmenliği	33	6.1
	Sosyal Bilgiler Öğretmenliği	102	19.0
	Türkçe Öğretmenliği	30	5.6

Tablo 1'e göre çalışma grubunun üçte ikisini kız öğrenciler ve neredeyse çeyreğini erkek öğrenciler oluşturmaktadır. Bununla birlikte, çalışma grubunun %70'inin kişisel bir bilgisayarı mevcutken %30'unun kişisel bir bilgisayarı bulunmamaktadır. Tablo 1'de katılımcıların %31.1'ini rehberlik ve psikolojik danışmanlık; %19'unu sosyal bilgiler öğretmenliği; %16.8'ini okul öncesi öğretmenliği; %9.1'ini matematik öğretmenliği öğrencileri oluşturmaktadır. Çalışma grubunun geri kalan kısmını ise beden eğitimi, fen bilgisi eğitimi, İngilizce öğretmenliği, müzik öğretmenliği, resim-iş öğretmenliği ve Türkçe öğretmenliği öğrencilerinin oluşturdukları görülmektedir.

Araştırmanın nitel kısmında, yüksek düzeyde DMF içerdiği belirlenen maddelerin yanlılık gösterip göstermediğini belirleyebilmek amacıyla uzman görüşlerine başvurulmuştur. Uzman grubunu, Eğitim Fakültesi'nin çeşitli bölümlerinde/anabilim dallarında görev yapan öğretim üyeleri ile doktora tamamlamış araştırmacılar oluşturmaktadır. Uzman grubu, Bilgisayar ve Öğretim Teknolojileri Bölümü'nden bir doktor araştırmacı başta olmak üzere, Eğitimde Ölçme ve Değerlendirme Anabilim Dalı'ndan iki doktor öğretim üyesi ve bir doktor araştırma görevlisi; Eğitim Yönetimi Anabilim Dalı'ndan bir doçent, yine Eğitimde Ölçme ve Değerlendirme alanında doktora tamamlamış bir Milli Eğitim Bakanlığı uzmanı ile rehberlik ve psikolojik danışmanlık alanında doktora tamamlamış bir araştırma görevlisi ve bir doktor öğretim üyesinden oluşmaktadır.

### 2.3. Veri toplama araçları ve süreci

Bu çalışmada, verinin toplanması aşamasında, Bahar (2014) tarafından geliştirilmiş, Bahar ve Asil (2018) tarafından revize edilmiş, E-Değerlendirmeye Yönelik Tutum Ölçeği (EDYTÖ) kullanılmıştır. 5'li Likert tipinde puanlanan ölçek, Sınav Özelliği, Bireysel/Fiziksel Özellikler, Teknik Faktörler ve Sosyal/Çevresel Faktörler boyutlarında yer alan 23 maddeden oluşmaktadır. AFA sonuçlarına göre dört boyut ve 23 madde ile toplam varyansın %59.5'i açıklanmakta; madde faktör yükleri birinci boyutta 0.62-0.81; ikinci boyutta 0.53-0.83; üçüncü boyutta 0.60-0.86 ve dördüncü boyutta ise 0.87-0.88 ranjında değer almaktadır. Ölçeğin tümü için hesaplanan Cronbach Alfa katsayısı 0.85'tir (Bahar, 2014). Bahar ve Asil (2018) çalışmasında, DFA sonucu gerçekleşen modifikasyon önerilerini alanyazın temelinde dikkate alarak modelde birtakım değişiklikler uygulamış ve ölçeğin dört faktörlü yapıya uygun olduğunu ortaya koymuştur ( $\chi^2 = 1262.53$ ;  $sd = 224$ ;  $RMSEA = 0.07$ ;  $CFI = 0.96$ ;  $TLI = 0.95$ ). EDYTÖ ile birlikte araştırmacı tarafından hazırlanmış demografik bilgiler formu da çalışma grubuna uygulanmıştır. Araştırmanın uzman görüşlerine başvuru nitel kısmında ise, yüksek düzeyde DMF içerdiği tespit edilen maddeler için açık uçlu sorulardan oluşan bir form yardımıyla uzman görüşleri toplanmıştır.

Veri toplama sürecinin başında ölçeğin sahibinden ve Etik Kurul'dan gerekli izinler alınmıştır. Nicel kısımda, EDYTÖ ve demografik bilgiler formu çevrim içi ortama aktarılmış, link öğrencilerle paylaşılmıştır. Çalışmaya katılımda gönüllülük esas alınmış ve katılımcılar arasından ölçeği iki kere dolduranlar belirlenerek veri setinden çıkarılmıştır. Araştırmanın nitel kısmında başvuru uzman



görüşleri ise yine araştırmaya katılmaya gönüllü uzmanlardan e-posta aracılığıyla toplanmıştır. Uzmanlardan görüşleri, uzman grubunun farklı alanlardan en az doktora sahibi araştırmacılardan oluşması sebebiyle, DMF kısaca açıklanarak ve madde bazında elde edilen sonuçlar hakkında bilgilendirme yapılarak elde edilmiştir.

#### 2.4. Verilerin analizi

Ölçme değişmezliği ve DMF analizlerine geçmeden önce, elde edilen veri setine ilişkin güvenilirlik hesaplanmıştır. Ölçeğin tümü için McDonald'ın Omega katsayısı 0.95'tir. Buna göre verinin güvenilirliğinin yüksek olduğu söylenebilir. Ardından araştırmanın bağımsız değişkenlerinin her alt grubunda yapı Doğrulayıcı Faktör Analizi (DFA) ile sınanmıştır. Bu noktada DFA öncesi veri setinin faktör analizi varsayımlarını sağlayıp sağlamadığı test edilmiştir. Veri setinde tek yönlü ve çok yönlü uç değerler z değerleri ve Mahalanobis uzaklığı yardımıyla incelenmiştir. Hesaplanan z değerlerine göre -3,3 aralığı dışında kalan 22 ve Mahalanobis uzaklığına göre kritik Kikare değeri ( $\alpha = 0.001$ ;  $sd = 22$ ) 48.268'yi aşan 15 birey analiz dışına çıkarılmıştır. Ardından hesaplanan KMO değerinin 0.94 ve Bartlett Küresellik Testi'nin manidar olduğu görülmüştür. Bununla birlikte maddeler arası ilişkiler anti-image korelasyon matrisi yardımıyla incelenmiş, beklenen şekilde yüksek düzeyde korelasyonların hesaplandığı görülerek DFA ile alt gruplarda ölçeğin yapısı test edilmiş ve sonuçlar aşağıda Tablo 2'de sunulmuştur.

**Tablo 2.**

*Alt gruplarda DFA Sonuçları*

	$\chi^2$	sd	$\chi^2/sd$	NFI	TLI	CFI	IFI	RMSEA	SRMR
<b>Kız</b>	948.46	223	4.25	0.94	0.95	0.95	0.96	0.09	0.07
<b>Erkek</b>	556.06	223	2.49	0.87	0.91	0.92	0.92	0.09	0.07
<b>PC var</b>	942.12	223	4.22	0.93	0.94	0.95	0.95	0.09	0.08
<b>PC yok</b>	430.85	223	1.93	0.93	0.96	0.96	0.96	0.07	0.07

Tablo 2'de  $\chi^2/sd$  sonucu hesaplanan değerler incelendiğinde "kız öğrenci" ve "PC var" alt gruplarında orta düzeyde uyum; "erkek" ve "PC yok" alt gruplarında ise mükemmel uyum olduğu söylenebilir. NFI değerlerine göre yalnızca "erkek" alt grubunda 0.90'a yakın bir değer çıkmış; diğer değerler ise iyi uyuma işaret etmektedir. TLI, CFI ve IFI'ye göre ise tüm alt gruplarda iyi düzeyde model veri uyumu yakalandığı görülmüştür. Hatalar çerçevesinde, RMSEA değerlerine göre "PC yok" alt grubunda iyi uyum ve diğer alt gruplarda ise zayıf uyum görülmekte iken; SRMR değerlerine göre tüm alt gruplarda model veri uyumunun iyi düzeyde olduğu söylenebilir (Kline, 2016). Bu bulgular ışığında alt gruplarda model veri uyumunun sağlandığı yorumu yapılabilir.

Birinci alt problemin çözümünde, Rosseel (2012) tarafından geliştirilmiş "Lavaan" ile Jorgensen vd. (2022) tarafından geliştirilmiş "SEMtools" paketleri yardımıyla R (2021) ortamında çoklu grup DFA uygulanarak cinsiyet ve kişisel bir bilgisayara sahip olma (PC) değişkenleri bakımından ölçme değişmezliği test edilmiştir. Değerlendirme kriteri olarak CFI ve TLI uyum indekslerinin fark ( $\Delta$ ) değerleri dikkate alınmıştır. Fark değerleri genellikle (-0.01, 0.01) aralığında değerlendirilse de bu değerlerin örneklem büyüklüğünden, kurulan modelden ve veri özelliklerinden etkilendiği ifade edilmektedir (Kline, 2016). Chen (2007) tarafından yürütülen bir simülasyon çalışması sonucunda, her grupta 300'den az sayıda gözlem bulunması ve gruplardaki gözlem sayılarının eşit olmaması durumunda  $\Delta CFI \leq .005$  kriterinin referans alınması önerilmiştir. Bununla birlikte  $\Delta RMSEA$  değerlerinin küçük örneklerde gerçek eş değeri reddetme eğiliminde olduğu belirtilmektedir (Chen, 2007). Alanyazında uyum değerlerinin örneklem büyüklüğüne duyarlılıklarına ilişkin kesin bir sonuca ulaşılamamış olması sebebiyle bu araştırmada  $\Delta CFI$  ve  $\Delta TLI \leq 0.005$  ölçüt alınmıştır.

İkinci alt problemin çözümünde, DMF gösteren maddelerin belirlenebilmesi amacıyla, tüm maddeler şüpheli madde ve şüpheli madde dışındaki tüm maddeler eşleştirme maddesi olarak analize dâhil edilmiştir. Analizlerde, SIBTEST programı ve POLY-SIBTEST yöntemi işe koşulmuştur (Shealy & Stout, 1993). POLY-SIBTEST, Shealy ve Stout (1993) tarafından geliştirilmiş Klasik Test Kuramı'na (KTK) dayalı

bir yöntem olmakla birlikte, çoklu puanlanan veya karma puanlama sistemine sahip ölçme araçlarında DMF belirleme amacıyla kullanılmamaktadır. POLY-SIBTEST ile DMF açısından incelenecek maddeler şüpheli madde (suspected items) ve eşleştirme maddeleri (matching items) olmak üzere iki alt teste ayrılmaktadır. Referans ve odak grubu oluşturan bireyler, toplam puana ve kategori sayısına göre eşleştirilir. Her bir gruptaki bireylerin, testle ölçülen özellik bakımından aynı yetenekte olduğu varsayılarak referans ve odak gruptaki bireylerin performansları her bir “çalışılan test” için K grupta karşılaştırılır (Henderson, 1999). POLY-SIBTEST sonuçlarının değerlendirilmesinde, Roussos ve Stout (1996) tarafından hazırlanan sınıflama dikkate alınmıştır. Bu sınıflamaya göre,  $|\beta| < 0.059$  ise DMF yok veya ihmal edilebilir düzeyde (A düzeyi);  $0.059 \leq |\beta| < 0.088$  ise orta düzeyde DMF (B düzeyi) ve  $|\beta| \geq 0.088$  ise yüksek düzeyde (C düzeyi) DMF yorumları yapılmaktadır (Roussos & Stout, 1996). Bununla birlikte Beta ( $\beta$ ) değerinin negatif olması odak grup lehine, pozitif olması ise referans grup lehine DMF varlığı anlamına gelmektedir.

Üçüncü alt problemde ölçek maddeleri, Meyer (2014) tarafından geliştirilmiş *jMetrik* programında MantelTest ile cinsiyet ve PC değişkenleri temelinde DMF bakımından analiz edilmiştir. MantelTest KTK’ya dayanan bir yöntem olmakla birlikte çoklu puanlanan maddeler için DMF belirlemede kullanılan yöntemler arasındadır. Program çıktılarında maddeler için hesaplanan  $\chi^2$  değerleri ve bunların anlamlılık düzeyleri ile etki büyüklükleri ve maddenin hangi grup lehine DMF içerdiği yer almaktadır. Ayrıca orta ve yüksek düzeyde etki büyüklüğüne sahip maddeler için eksi ve artı işaretleri incelenerek maddenin hangi grup lehine DMF gösterdiği kolaylıkla yorumlanabilmektedir. Eğer madde eksi işaretliyse referans grup lehine; artı işaretliyse odak grup lehine DMF içerdiği yorumunda bulunmaktadır (Meyer, 2014). Araştırmanın nitel kısmını oluşturan dördüncü alt problemin çözümünde, POLY-SIBTEST ve MantelTest sonucu yüksek düzeyde DMF içerdiği tespit edilen maddeler için uzman görüşleri alınmış; bu maddelerin yanlılık gösterip göstermediği uzman görüşleri temelinde yorumlanmıştır.

## 2.5. Araştırmanın etik izni

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

### Etik kurul izin bilgileri

Etik değerlendirmeyi yapan kurul adı: Aksaray Üniversitesi İnsan Araştırmaları Etik Kurulu

Etik değerlendirme kararının tarihi: 21.12.2021

Etik değerlendirme belgesi sayı numarası: 2021/08-50

## 3. BULGULAR

Birinci alt problemde, e-Değerlendirmeye Yönelik Tutum Ölçeği’nin (EDYTÖ) cinsiyet ve PC değişkenleri temelinde ölçme değişmezliğini ne düzeyde sağladığı incelenmiştir. Bu amaçla yürütülen analiz sonucu cinsiyete ve PC’ye göre elde edilen sonuçlar aşağıda Tablo 3’te özetlenmiştir. Tablo 3’te yer alan  $\Delta$  CFI ve  $\Delta$  TLI değerleri incelendiğinde bu değerlerin -0.005, 0.005 aralığının dışında olduğu görülmektedir. Buna göre, EDYTÖ’nün cinsiyet ve PC değişkenleri temelinde metrik değişmezliği sağlamadığı söylenebilir. Bu bulguya dayanarak araştırmanın ikinci alt problemi olan DMF analizine geçilmiştir.

İkinci alt problemin çözümünde, EDYTÖ’de yer alan maddelerin tümü şüpheli ve madde dışındakiler eşleştirme maddesi olmak üzere, cinsiyet ve PC değişkenlerine göre maddelerin DMF içerip içermediği belirlenmiştir. Referans grup erkek öğrenciler ve odak grup kız öğrenciler olmak üzere cinsiyet temelinde; referans grup PC’si olan ve odak grup PC’si olmayan öğrenciler olmak üzere PC değişkeni temelinde yürütülen POLY-SIBTEST sonuçları Ek-1’de sunulmuştur.

**Tablo 3.***Cinsiyete ve PC'ye Göre Değişmezlik Sonuçları*

		Kikare	sd	p	CFI	TLI	RMSEA	SRMR	Δ CFI	Δ TLI
<b>Cinsiyet</b>	Yapısal	1034.710	448	0.00	0.81	0.79	0.07	0.07		
	Metrik	832.950	467	0.00	0.88	0.87	0.06	0.08	0.07	0.08
	Skalar	862.403	486	0.00	0.88	0.88	0.06	0.08	-0.03	0.00
	Katı	885.860	509	0.00	0.88	0.88	0.06	0.08	0.00	0.00
<b>PC</b>	Yapısal	1016.712	448	0.00	0.82	0.80	0.07	0.07		
	Metrik	792.485	467	0.00	0.90	0.89	0.05	0.07	0.08	0.09
	Skalar	824.519	486	0.00	0.90	0.89	0.05	0.07	-0.00	0.00
	Katı	839.130	509	0.00	0.90	0.90	0.05	0.07	0.00	0.00

Ek-1'de cinsiyete göre DMF bulguları incelendiğinde 5, 6, 13, 14, 17 ve 19 numaralı maddelerin  $\beta$  değerlerinin anlamlı olduğu görülmektedir ( $p < 0.05$ ). Buna göre bu maddeler cinsiyet değişkenine göre DMF göstermektedir. Hesaplanan  $\beta$  katsayılarına göre, tüm değerlerin 0.088'den büyük olması sebebiyle, bu maddelerin yüksek düzeyde (C) DMF içerdiği söylenebilir. Pozitif  $\beta$  değerine sahip 5, 6, 17, 19 numaralı maddelerin referans grup olan erkek öğrenciler lehine; negatif  $\beta$  değerine sahip 13 ve 14 numaralı maddelerin de odak grup olan kız öğrenciler lehine yüksek düzeyde DMF gösterdiği tespit edilmiştir.

Ek-1'deki PC değişkenine göre DMF bulguları incelendiğinde ise, yalnızca 16 numaralı maddenin  $\beta$  değerinin manidar olduğu ( $p < 0.05$ ) ve diğer maddelerin  $\beta$  değerlerinin anlamlı olmadığı ( $p > 0.05$ ) görülmüştür. Buna göre Madde 16'nın kişisel bir bilgisayara sahip olma değişkeni temelinde DMF gösterdiği yorumu yapılabilir. Hesaplanan  $\beta$  değerinin -0.147 olması ise bu maddenin odak grup olan PC'si olmayan öğrenciler lehine ve yüksek düzeyde (C) DMF içerdiğine işaret etmektedir.

Üçüncü alt problemin çözümünde, referans grup erkek ve odak grup kız öğrenciler olmak üzere cinsiyet temelinde; referans grup PC'si olmayan ve odak grup PC'si olan öğrenciler olmak üzere PC değişkeni temelinde yürütülen MantelTest sonuçları Ek-2'de yer almaktadır. Ek-2'de yer alan bulgulara göre, 12, 13 ve 20 numaralı maddeler orta düzeyde (B) DMF içerirken; Madde 6'nın referans grup olan erkek öğrenciler lehine yüksek düzeyde (C) DMF gösterdiği söylenebilir. PC değişkenine göre ise tüm maddelerin ihmal edilebilir düzeyde (A) DMF gösterdiği görülmektedir. Dördüncü alt probleme geçmeden, her iki yöntemle yüksek düzeyde DMF içerdiği belirlenen maddelere aşağıda Tablo 4'te yer verilmiştir.

**Tablo 4.***Yüksek Düzeyde DMF İçerdiği Belirlenen Maddeler*

Madde	Avantaj Sağladığı Grup
Madde 5. İnternet üzerinden yapılan sınav esnasında bağlantının kopmasından korkarım.	Erkek öğrenciler
Madde 6. İnternet üzerinden yapılan sınavda gözlerim daha çok yorulur.*	
Madde 17. İnternet üzerinden yapılan sınavlar daha güvenilir sonuçlar verir.	
Madde 19. İnternet üzerinden yapılan sınavlarda hatalar daha az olur.	
Madde 13. İnternet üzerinden yapılan sınavda daha iyi konsantre olabilirim.	Kız öğrenciler
Madde 14. İnternet üzerinden yapılan sınavlar benim kişilik özelliklerime daha uygundur.	
Madde 16. İnternet üzerinden yapılan sınav bütün becerilerimi ölçebilir.	PC'si olmayanlar

\*Hem SIBTEST hem MantelTest sonucu yüksek düzeyde DMF içeren maddedir.

Beşinci alt problemde, yüksek düzeyde DMF içerdiği tespit edilen tüm maddelere ilişkin uzman görüşleri incelenmiş; DMF'nin madde etkisinden mi yoksa yanlılıktan mı kaynaklandığı ortaya koyulmaya çalışılmıştır. Aşağıda öncelikle erkek öğrenciler lehine; ardından kız öğrenciler lehine ve son olarak PC



sahibi olmayan grubun lehine yüksek düzeyde DMF içerdiği belirlenen maddelere yönelik uzman görüşleri, uzmanlar U1, U2 şeklinde adlandırılarak ele alınmıştır.

### Erkek Öğrenciler Lehine DMF Belirlenen Maddelere Yönelik Uzman Görüşleri

#### *Madde 5. İnternet üzerinden yapılan sınav esnasında bağlantının kopmasından korkarım.*

Madde 5, e-değerlendirmeye yönelik olumsuz bir tutum ifadesidir ve bu maddenin erkek öğrenciler lehine DMF göstermesi, aynı tutum düzeyindeki kız ve erkek grupların maddeye verdiği cevapların erkekler lehine farklılaştığı anlamına gelmektedir. Maddenin olumsuz bir tutum ifadesi olması ve ters puanlanması sebebiyle, Madde 5'in dâhil edilmediği eşleştirme testinde erkeklerle aynı tutum düzeyindeki kız öğrencilerin korkusunun daha yüksek olduğu söylenebilir. Bu bulguya ilişkin uzman görüşleri aşağıda yer almaktadır. Uzman görüşleri incelendiğinde, uzmanların bilgisayar/teknoloji kullanımının erkekler arasında daha yaygın olduğu konusunda birleştikleri, kız öğrencilerin erkek öğrencilere kıyasla daha yüksek kaygı düzeyine sahip olduğunu düşündükleri, bununla birlikte erkek öğrencilerin bağlantı problemini bir stres kaynağı olarak algılamayabilecekleri veya algılasalar bile bu problemi hızlıca çözebileceklerine inançları ifade edilmiştir. Buna göre Madde 5'in yanlılık içermediğini, erkekler lehine ortaya çıkan bu farklılığın madde etkisinden kaynaklandığını söylemek mümkündür.

**U1:** “Bağlantının kopması ile ilgili kaygının temelinde çevrim içi sistemlerde bir veri akışının olduğu ve bu veri akışında sorun yaşandığında cevaplarının sisteme kayıt olmayacağı bilgisi vardır. Bu tarz kopmaların sonuçlarının daha çok çevrim içi oyunlarda görüldüğü söylenebilir. Çevrim içi oyun çalışmalarında da erkeklerin daha fazla oyun oynama eğilimi olduğu ifade edilmektedir. Buna bağlı olarak bu maddedeki yanlılığın çevrim içi oyunlara bağlı olarak erkeklerin lehine çıktığı düşünülmektedir.”

**U2:** “Erkek öğrenciler teknoloji ile daha ilgili olabileceklerinden onların korkuların kadın öğrencilerin korkularından daha azdır. Bundan dolayı DMF içerir bu madde.”

**U3:** “Maddede kullanılan sözcük sebebi ile erkekler lehine DMF tespit edilmiş olabilir. Bağlantının kopmasından korkarım ifadesindeki korkarım kelimesi kız öğrencilerin kaygı durumları ile ilişkilendirilebilir. Yapılan araştırmalarda genellikle kız öğrencilerin kaygı puanları erkek öğrencilerden daha yüksek bulunduğu için erkek öğrenciler lehine DMF tespit edilmiş olabilir.”

**U4:** “Kız öğrencilerin teknoloji ile ilişkisi erkek öğrencilere göre daha olumsuz olabilir. Diğer bir neden kız öğrencilerde sınav kaygısının daha yüksek olması. Erkek öğrenciler bu konuda daha rahat davranıyor olabilir.”

**U5:** “Madde üzerinde ve maddenin ortaya çıkartabileceği anlam bakımından internet bağlantısının olası nedenlerinin her ne kadar sistemselsel olsa da öğrenci ile alakalı (internet kopma, wifi bağlanmama) değişkenler olabileceğinden hareketle ortaya çıkan istatistiklerin farklılaştığı dolayısıyla özellikle korku gibi birey odaklı bir maddenin öğrenci yani birey lehinde çıkmış olma olasılığı doğal bir sonuç gibi görünmektedir. Cinsiyet bakımından bakıldığında ise erkeklerin genel toplumsal araştırmalarda psikolojik özelliklerinin kızlara göre standart ölçme araçlarında düşük olduğu genel ampirik bulguları mevcuttur.”

**U6:** “Burada erkek öğrencilerin lehine ile daha yüksek korkuya sahip olduğu kastediliyor sanırım; negatif puanlama tam tersi anlam da çıkabilir. Hangi anlam olduğundan bağımsız olarak bu maddenin cinsiyete göre bir yanlılık içermediği kanaatindeyim. Maddeye verilen cevapların gerçek tepkiyi gösterdiğini düşünüyorum.”

**U7:** “Bu maddede erkeklerin lehine sonuç elde edilmesi erkeklerin bilgisayar ve internet konusunda kızlara göre daha kendilerini yeterli hissetmeleri ve bir problem yaşasalar bile bir veya birden fazla çözümü daha hızlı ve kolay bir şekilde ortaya koyacaklarına ilişkin inançları olabilir.”

**U8:** “Erkeklerin kadınlarla kıyaslandığında bilişim teknolojileri alanında bilgi ve becerilerinin daha fazla olması bunun nedeni olabilir. Ayrıca bağlantı koptuğu zaman bununla baş edebileceğine dair daha fazla öz güvenleri olabilir. Erkeklerin bilişim teknolojilerinin kullanımına yönelik öz yeterlik algıları kadınlara göre daha yüksektir. Bu konuda yapılan çalışmalar da buna kanıttır (Kartal, Temelli & Şahin, 2018). İkinci olarak sınav esnasında bağlantının

*kopması bir stres faktörüdür. Kadınlar ve erkeklerin stresi algılama biçimleri, strese verdikleri tepkiler ile stresle baş etme şekilleri birbirinden farklıdır. Yani erkekler bağlantı kopmasını bir stres, problem olarak algılamıyor olabilir."*

**Madde 6. İnternet üzerinden yapılan sınavda gözlerim daha çok yorulur.**

Madde 6'nın, hem POLY-SIBTEST hem de MantelTest sonucu yüksek düzeyde DMF içerdiği belirlenmiştir. Maddenin e-değerlendirmeye yönelik olumsuz bir tutum ifadesi olması ve erkek öğrenciler lehine DMF göstermesi, aynı tutum düzeyindeki kız ve erkek grupların maddeye verdiği cevapların erkekler lehine farklılaştığı anlamına gelmektedir. Bu bulguya ilişkin uzman görüşleri aşağıda sunulmuştur. Uzman görüşleri incelendiğinde, uzmanların maddeye ilişkin görüşlerinin farklılaştığı söylenebilir. Uzmanlardan biri maddenin sorgulanması gerektiğini ifade ederken; beş uzmanın bu durumun beklenen bir sonuç olduğunu düşündüğü ve ikisinin ise nötr kaldıkları görülmektedir. Bu açıdan bakıldığında maddede ortaya çıkan bu farklılığın alanyazın eşliğinde detaylı bir şekilde incelenmesinin gerektiği düşünülmüştür.

**U1:** "Ekran üzerinden okumaya yönelik göz izleme çalışmaları bulunmaktadır. Bu çalışmalarda cinsiyete yönelik farklılıklar olduğu da raporlanmaktadır. Ancak bu farklılığın nedeni yapısal bir nedenden mi yoksa farklı deneyimlerden kaynaklı olup olmadığı ifade edilmemektedir. (Örnek bir çalışma için: Zhan, Wu, Mei, Fong, Huang, & Shao, 2019). Ayrıca sınav kaygısı ile ilgili cinsiyete yönelik farklılıklar var ise bu çalışmalar yol gösterebilir."

**U2:** "Yukarıda (Madde 5 için) bahsedilen gerekçe olabildiği gibi erkekler daha çok alışık olabilirler internet üzerinde yapılan sınavlara. Bu da bir DMF göstergesi olabilir."

**U3:** "Cinsiyete göre öğrencilerin internet kullanım süresi açısından, aralarında istatistiksel olarak anlamlı farklılık bulunan çalışmalar mevcut (Kelleci vd., 2009). Bu çalışmalarda da olduğu gibi erkek öğrenciler internette kız öğrencilere göre daha fazla zaman geçirmekte, bu bağlamda maddedeki DMF'nin erkekler lehine olması, erkek öğrencilerin daha fazla aşına oldukları bir platformda bedensel ya da fizyolojik olarak daha rahat hissetmelerinden kaynaklanabilir."

**U4:** "Çok anlamı gelmedi açıkçası ama kız öğrencilerin sınavlara ya da ders ortamına daha iyi konsantre olması ile ilişkilendirilebilir."

**U5:** "Göz yorulması gerekçelerinin tıbbi veya biyolojik nedenleri ve ekran bakma ile ilişkisinin bulunduğu teorik veya ampirik araştırmalarla nedeni üzerinde yorum yapmanın uygun olacağı ve bu konuda bilgimin neden geliştirecek kadar yeterli olduğunu düşünmüyorum."

**U6:** "Maddede yanlılık oluşturan bir durum olmadığını düşünüyorum."

**U7:** "Bu maddeyi en sona bıraktım değerlendirmede çünkü bir türlü farkın çıkmasını kafamda oturtamadım. Bu doğrudan fizyolojik bir durum, cinsiyetten bağımsız olmasını beklerdim ve hala da bekliyorum. O nedenle bu farkın sorgulanabileceğini düşünüyorum."

**U8:** "Erkeklerin bilişim teknolojileri daha fazla ilgili olmaları ve uzun süreli vakit geçirmelerinden kaynaklı alışık olmaları bunun bir nedeni olabilir. Ayrıca erkeklerin beden duyumlarına ilişkin farkındalık düzeyleri kadınlardan daha düşük olabilir."

**Madde 17. İnternet üzerinden yapılan sınavlar daha güvenilir sonuçlar verir.**

Madde 17, e-değerlendirmeye yönelik olumlu bir tutum ifadesidir ve bu maddenin erkek öğrenciler lehine yüksek düzeyde DMF göstermesi, aynı tutum düzeyindeki kız ve erkek grupların maddeye verdiği cevapların erkekler lehine farklılaştığı anlamına gelmektedir. Madde 17'nin dâhil edilmediği eşleştirme testinde kızlarla aynı tutum düzeyindeki erkek öğrencilerin güvenilirliğe yönelik tutumlarının daha yüksek olduğu söylenebilir. Bu bulguya ilişkin uzman görüşleri aşağıda yer almaktadır. Uzman görüşleri incelendiğinde, Madde 17'de ortaya çıkan bu farkı da madde etkisine dayandırmak mümkündür.

**U1:** "Tutum deneyimler sonucu şekillenmektedir. Bu yüzden çevrim içi oyun ya da programlama deneyimleri erkeklerde daha fazla olduğu alanyazında ifade edildiği için, bu maddede de bir eğilime neden olmuş olabilir. Bu

deneyimlere sahip kişiler arka planda çalışan algoritmadan haberdar oldukları için insandan kaynaklı hatalardan daha aza indiğini biliyorlardır.”

**U2:** “Güvenilir olmasının sebebi erkeklerden beklenen davranışların kadın öğrencilerinden de beklenmesidir. Yani olması gereken davranışı kadınların göstermemesi gözlemlenecek olan davranışı ölçmede eksi duruma getirebilir.”

**U3:** “Öğrencilerin e-sınavların güvenilirliği konusundaki görüşleri arasındaki farkın sebebi yine erkek öğrencilerinin internet platformuna daha aşina olmasına (daha fazla vakit geçirmesi) bağlayabilirim (platformu ne kadar iyi tanır ve bilirse o kadar hatasız kullanabilir). Yapılan çalışmalarda erkek öğrencilerin kız öğrencilere göre internet platformlarında yapılan sınav uygulama ya da değerlendirmelerin daha fazla yaygınlaştırılmasını isteyen tarafta olduğu görülüyor (Başol vd., 2017).”

**U4:** “Erkek öğrencilerin teknoloji ile daha ilişkili olmasından kaynaklanabilir.”

**U5:** “Güvenilirliğin temel anlamda hatalardan arınıklığı ifade ettiğini düşündüğümüzde kişisel olarak özellikle üniversitemizin bize sunduğu imkânlar kapsamında uzaktan yaptığım sınavların güvenilirliğinin sınav güvenliğinin sağlayamadığım gerekçesiyle düşük olduğunu özellikle kopya vb çekerek sınava hata karışma ihtimalinin oldukça yüksek olduğu kanısındayım. Bu nedenle cinsiyet üzerinde özellikle başarı olarak kız öğrencilere göre düşük olan erkek öğrenciler lehine bu şekilde bir farklılığın oluşmasını normal olduğunu düşünüyorum.”

**U6:** “Maddede yanlılık oluşturan bir durum olmadığını düşünüyorum.”

**U7:** “Bu madde de erkeklerin daha yüksek puan almasının nedeni, erkeklerin bilgisayara dünyasına daha fazla hâkim olması, kızlara göre internet ve bilgisayarda daha fazla zaman geçirmeleri, bu ortamı daha iyi bilmeleri olabilir.”

**U8:** “Bilişim teknolojileri kullanılarak yapılan sınavlarda yüz yüze yapılan sınavlara göre öğrenciden ve öğretmenden kaynaklı hata daha az olabilir. Gözlemlerime göre erkek öğrencilerin yüz yüze yapılan sınavlarda kodlama yaparken kaydırma yapma veya optik formu yanlış şekilde işaretleme olasılığı daha fazla olduğundan, online sınavların daha güvenilir olduğunu düşünebilirler. Uzaktan ölçme ve uzaktan ders niteliğine yönelik algılar arasında erkek öğrencilerin kadın öğrencilerden daha olumlu algıya sahip olduğu bulunmuştur (Yıldırım, 2021).”

#### **Madde 19. İnternet üzerinden yapılan sınavlarda hatalar daha az olur.**

Madde 19, e-değerlendirmeye yönelik olumlu bir tutum ifadesidir ve bu maddenin erkek öğrenciler lehine DMF göstermesi, aynı tutum düzeyindeki kız ve erkek grupların maddeye verdiği cevapların erkekler lehine farklılaştığı anlamına gelmektedir. Madde 19’un dâhil edilmediği eşleştirme testinde kızlarla aynı tutum düzeyindeki erkek öğrencilerin hatalara/güvenirliliğe ilişkin tutumlarının daha yüksek olduğu söylenebilir. Bu bulguya ilişkin uzman görüşleri aşağıda sıralanmıştır. Uzman görüşleri dikkate alındığında, Madde 17 ile benzer şekilde, bu maddede de madde etkisinden kaynaklanan bir farklılığın olduğuna karar verilmiştir.

**U1:** “Buradaki ifade aslında çok net değil. Çünkü soruların sisteme aktarılması mı, soruya cevap verme mi, yoksa puanlama aşaması mı kastediliyor anlaşılmamaktadır. Eğer puanlama aşaması ise bir önceki madde için yaptığım açıklama bu madde için de geçerlidir.”

**U2:** “Daha alışık olmalarından kaynaklı olarak bu madde tamamen erkek öğrenciler lehinedir.”

**U3:** “Madde 17 için verdiğim cevap geçerli.”

**U4:** “Erkek öğrencilerin teknoloji ile daha ilişkili olmasından kaynaklanabilir.”

**U5:** “Madde 17 ile aynı gerekçe olduğunu düşünüyorum.”

**U6:** “Maddede yanlılık oluşturan bir durum olmadığını düşünüyorum.”

**U7:** “Bu madde için de yorumum madde 17 ile özdeştir.”

**U8:** “Erkek öğrencilerin yine bilişim teknolojileri ile ilgili bilgi ve becerilerinin daha yüksek olmasından kaynaklı online sınavlara yönelik algılarının daha olumlu olması ile ilgilidir. Şöyle ki, internet üzerinden yapılan sınavlar çoğunlukla çoktan seçmeli veya kısa cevaplı olmaktadır. Doğru-yanlış yanıtların belirlenmesi bilgisayar aracılığı ile yapıldığı için sınavla ilgili hata oranı azalmaktadır. Üstelik yüz yüze yapılan sınavlarda öğrencilerin optik formda kaydırma yapması vb. durumlar öğrencinin daha düşük puan alması ile sonuçlanmaktadır. Bu tür durumları erkeklerin yaşama olasılığı kadınlara göre daha fazladır. Bu nedenle erkekler lehine bir sonuç çıkmış olabilir.”

### **Kız Öğrenciler Lehine DMF Belirlenen Maddelere Yönelik Uzman Görüşleri**

#### **Madde 13. İnternet üzerinden yapılan sınavda daha iyi konsantre olabilirim.**

Madde 13, e-değerlendirmeye yönelik olumlu bir tutum ifadesidir ve bu maddenin kız öğrenciler lehine DMF göstermesi, aynı tutum düzeyindeki kız ve erkek grupların maddeye verdiği cevapların kızlar lehine farklılaştığı anlamına gelmektedir. Madde 13’ün dâhil edilmediği eşleştirme testinde erkeklerle aynı tutum düzeyindeki kız öğrencilerin odaklanmaya ilişkin tutumlarının daha yüksek olduğu söylenebilir. Bu bulguya ilişkin uzman görüşleri aşağıda sunulmuştur. Uzman görüşlerine göre, Madde 13’teki farklılığın madde etkisinden kaynaklandığını söylemek mümkündür. Uzmanlara göre, görsel zekâ, başarı odaklı motivasyon ve sınav kaygısı temelinde düşünüldüğünde kız öğrencilerin ekran başında daha iyi konsantre olması beklenen bir sonuçtur.

**U1:** “Daha önce ifade ettiğim sınav kaygısı ile ilgili cinsiyete yönelik farklılıklar var ise bu çalışmalar yol gösterebilir.”

**U2:** “Kız öğrencilerin görsel zekâları erkek öğrencilere kıyasla daha yüksek düzeyde olduğundan bilgisayar ortamına alışkın kız öğrenciler erkek öğrencilerden daha fazla başarı gösterebilir.”

**U3:** “Kız öğrencilerin kaygı puanlarının yapılan çalışmalarda erkek öğrencilere göre daha yüksek olduğu belirtilmişti, sınıf ortamında yapılan sınavlarda kaygının daha fazla olması beklenen bir durumdur. Dolayısı ile özellikle kız öğrencilerde kaygı, telaş ya da heyecanın artmasına bağlı olarak konsantrasyon kaybı yaşanabilir. Bunun yerine kullanılan e-sınavlar kız öğrencilerin yüz yüze değerlendirmelere göre daha rahat bir ortam sağlayabilir ve öğrenciler daha konsantre olabilirler.”

**U4:** “Kız öğrencilerin sınavlara ya da ders ortamına daha iyi konsantre olması ile ilişkilendirilebilir.”

**U5:** “Madde 14’le benzer bir neden olduğunu düşünüyorum.”

**U6:** “Maddede yanlılık oluşturan bir durum olmadığını düşünüyorum.”

**U7:** “Burada farkın kızların lehine olması kendi kendilerine kaldıkları bir ortamda daha iyi konsantre olmalarından kaynaklanabilir. Örneğin kopya isteyen muhtemel grup üyeleri yoktur.”

**U8:** “Kız öğrencilerin başarı odaklı motivasyon düzeyleri erkeklere daha yüksektir. Yüz yüze sınav sırasında motivasyon düzeylerini bozacak bazı faktörler olabilir. Örneğin sıranın sallanması, gıcirtısı, diğer öğrencilerin hareketleri, sesleri, gözetmenin sesi, hareketi vb. Online sınavlarda ise bu faktörler elimine edilmektedir. Dolayısıyla daha fazla konsantre olduklarını düşünüyor olabilirler. Erkeklerin bu faktörlerden etkilenme olasılıkları daha düşük olabilir. Kadınlar çevresel uyaranlara daha duyarlı olabilirler.”

#### **Madde 14. İnternet üzerinden yapılan sınavlar benim kişilik özelliklerime daha uygundur.**

Madde 14, e-değerlendirmeye yönelik olumlu bir tutum ifadesidir ve bu maddenin kız öğrenciler lehine DMF göstermesi, aynı tutum düzeyindeki kız ve erkek grupların maddeye verdiği cevapların kızlar lehine farklılaştığı anlamına gelmektedir. Madde 14’ün dâhil edilmediği eşleştirme testinde erkeklerle aynı tutum düzeyindeki kız öğrencilerin uygunluğa ilişkin tutumlarının daha yüksek olduğu söylenebilir. Bu bulguya ilişkin uzman görüşleri aşağıda sunulmuştur. Uzman görüşlerine göre, Madde 13’e benzer şekilde, görsel zekâ, başarı odaklı motivasyon ve sınav kaygısı temelinde düşünüldüğünde kız öğrencilerin ekran başında daha iyi konsantre olması beklenen bir sonuçtur. Bu bağlamda Madde 14’teki farklılığın da madde etkisinden kaynaklandığı söylenebilir.

**U1:** “Daha önce ifade ettiğim sınav kaygısı ile ilgili cinsiyete yönelik farklılıklar var ise bu çalışmalar yol gösterebilir.”



U2: "Görsel zekâdan kaynaklı olabileceğini düşünüyorum."

U3: "Bir önceki maddede belirttiğim gibi kız öğrenciler internet ortamında yüz yüze ortama göre kendilerini daha rahat hissedebilirler, kaygıları azalır, daha konsantre, daha sakin olabilirler. Dolayısı ile internet üzerinden yapılan sınavların kendilerinin kişilik özelliklerine daha uygun olduğunu düşünebilirler."

U4: "Sınav ortamı erkek öğrenciler için herhangi bir fark yaratmıyor olabilir."

U5: "Özellikle büyük çaplı, ülkelerin karşılaştırıldığı OECD, TIMMS vb. başarı ve kişinin diğer özellikleri odaklı karşılaştırmalı büyük verilerin bulunduğu sınavlarda kız öğrenciler lehine özellikle "kişilik" değişkeninde sonuçların varlığı bilinmektedir. Bu bağlamda çok dolaylı da olsa ölçütü sağladığı yönünde bir görüş geliştirmeye dayanak sağlanmış olduğumuzu düşünmekteyim."

U6: "Maddede yanlılık oluşturan bir durum olmadığını düşünüyorum."

U7: "Bu madde için de Madde 13 için yaptığım yorumu düşünmekteyim."

U8: "Bununla ilgili net bir görüşüm yok maalesef. Belki bireysel çalışmaya ve konsantre olmaya yönelik istek ve eğilimleri daha fazla olabilir. Çünkü kız öğrencilerin başarı odaklı motivasyon düzeyleri daha yüksektir."

### **Kişisel Bir Bilgisayarı Olmayan Öğrenciler Lehine DMF Belirlenen Maddelere Yönelik Uzman Görüşleri**

#### **Madde 16. İnternet üzerinden yapılan sınav bütün becerilerimi ölçebilir.**

Madde 16, e-değerlendirmeye yönelik olumlu bir tutum ifadesidir ve bu maddenin kişisel bir bilgisayara (PC) sahip olmayan öğrenciler lehine DMF göstermesi, aynı tutum düzeyindeki PC'ye sahip olan ve olmayan grupların maddeye verdiği cevapların PC'ye sahip olmayanlar lehine farklılaştığı anlamına gelmektedir. Madde 16'nın dâhil edilmediği eşleştirme testinde PC'ye sahip olanlarla aynı tutum düzeyindeki PC'ye sahip olmayan öğrencilerin bu madde temelinde tutumlarının daha yüksek olduğu söylenebilir. Bu bulguya ilişkin uzman görüşleri ise aşağıda yer almaktadır. Uzman görüşleri dikkate alındığında, PC sahibi olmayan grubun bu araca fazla anlam yüklemesi veya bu aracı pek bilmemesi konusunda fikirlerini ifade ettikleri görülmektedir. Buna dayanarak maddedeki farkın madde etkisinden kaynaklığı söylenebilir.

U1: "Araca aşına olma durumu kişinin kendini ifade etme durumunu etkilemektedir. Bu yüzden farklı bir araca alışma süreci söz konusu olduğu için bu öğrenciler bütün becerilerini ortaya koyamayacaklarını düşünüyor olabilir."

U2: "Kendi bilgisayarı olan bireylerin bilgisayar hakkında daha fazla bilgiye sahip olmasından kaynaklı olabilir."

U3: "Beklenen sonuç aleyhine çıkması iken lehine DMF olması şaşırtıcı, olası bir sebep söyleyemiyorum."

U4: "Kendisinin sahip olmadığı bir araca fazla anlam yüklemekten kaynaklanabilir."

U5: "Mantık örgüsü içerisinde düşünüldüğünde doğal bir sonuç olduğunu. Dolayısıyla asıl farklılığa neden olan değişkenin "hiç kişisel bilgisayara sahip olmama" veya "sahip olunan bilgisayara sahip olma ayı/yılı" değişkeni bakımından test edilmesi önerilir."

U6: "Bilgisayar sahibi olmayan öğrencilerin bilgisayar ve internete yüklediği anlamın daha fazla olabileceğini düşünüyorum ancak bu biraz kişisel bir düşünce. Bilimsel bir karşılığı var mı emin değilim. Ancak yine de maddede yanlılık oluşturan bir durum olmadığını düşünüyorum."

U7: "Bu sonuç bilgisayara sahip olmayan öğrencilerin yaşadıkları sorunu bir eksikliğe ki -o bilgisayarlarının olmaması- bağlamaları açıklanabilir. Yani bilgisayarları olsa yaşadıkları birçok sorunu aşacaklarına ve hatta dünyaya önemli katkılar sağlayacaklarına ilişkin alguları olabilir. Diğer yandan bilgisayarı olanlar bilgisayarın hiç de böyle bir işlevi olmadığını fark etmiş olabilirler."



U8: “Aslında böyle bir madde üzerinde tersi bir sonuç beklerdim. Kişisel bilgisayara sahip olmayanlar bilişim teknolojilerinin özellikleri hakkında daha yanlı ve abartılı bir öngörüye sahip olabilirler.”

#### 4. TARTIŞMA ve SONUÇ

Bu araştırmada, Bahar (2014) tarafından geliştirilmiş, Bahar ve Asil (2018) tarafından revize edilmiş e-Değerlendirmeye Yönelik Tutum Ölçeği'nin (EDYTÖ), cinsiyet değişkenine göre kız ve erkekler öğrenciler ile PC değişkenine göre kişisel bir bilgisayara sahip olan ve olmayan öğrenciler arasında ölçme değişmezliğinin sağlanamadığı sonucuna ulaşılmıştır. Metrik değişmezliğin sağlanamaması sebebiyle, ölçek maddelerinin DMF içerip içermediği incelenmiştir. Ölçek maddelerinin şüpheli madde ve madde dışındaki tüm maddelerin eşleştirme maddesi olması durumunda, cinsiyet ve PC değişkenleri temelinde yürütülen POLY-SIBTEST sonuçlarında; 5, 6, 17 ve 19 numaralı maddelerin erkek öğrenciler lehine yüksek düzeyde; 13 ve 14 numaralı maddelerin kız öğrenciler lehine yüksek düzeyde; 16 numaralı maddenin kişisel bir bilgisayara sahip olmayan öğrenciler lehine yüksek düzeyde DMF içerdiği tespit edilmiştir. MantelTest sonucuna göre ise yalnızca Madde 6'nın erkek öğrenciler lehine yüksek düzeyde DMF gösterdiği sonucuna ulaşılmıştır. Analiz sonucu yüksek düzeyde DMF içerdiği belirlenen tüm maddelere yönelik uzman görüşleri incelenmiş olup Madde 6 dışındaki maddelerin yanlılık göstermediği; DMF analizi sonucu ortaya çıkan bu farklılıkların madde etkisinden kaynaklandığı yorumuna gidilmiştir. Aşağıda, yüksek düzeyde DMF tespit edilerek uzman görüşüne başvurulmuş her bir madde alanyazın temelinde ele alınmış ve yorumlanmıştır.

“Madde 5. İnternette uygulanan sınavlarda bağlantının kopmasından korkarım.” ifadesinde kız öğrencilerin aleyhine yüksek düzeyde DMF belirlenmiştir. Bir başka deyişle, erkek öğrencilerle aynı tutum düzeyindeki kız öğrencilerin korkusunun daha yüksek olduğu görülmüştür. Uzman görüşlerine göre, kız öğrencilerin erkek öğrencilere kıyasla daha yüksek kaygı düzeyine sahip olması bu durumun olası nedenleri arasında yer almaktadır. Alanyazın incelendiğinde de, sınavlarda kız öğrencilerin daha yüksek kaygı yaşadıkları sonucuna ulaşılmıştır (Guida & Ludlow, 1989; Güler & Çakır, 2013; Kavakçı vd., 2011; McLean & Anderson 2009; Uzun Özer & Topkaya, 2011; Yıldız, 2007). Uzman görüşleri ve alanyazın birlikte değerlendirildiğinde Madde 5'te belirlenen DMF'nin madde etkisinden kaynaklanabileceği sonucuna varılmıştır.

“Madde 6. İnternet üzerinden yapılan sınavda gözlerim daha çok yorulur.” ifadesinde her iki yöntemle kız öğrencilerin aleyhine yüksek düzeyde DMF belirlenmiş olup tüm uzman görüşleri üzerinden fikir birliğine varılamaması dikkat çeken bir sonuç olmuştur. Benzer görüş belirten uzmanlar ise bilgisayar kullanımının erkekler arasında daha yaygın olduğunu ve bu nedenle e-değerlendirme uygulamalarında erkeklerin kendilerini daha rahat hissedeceğini, ayrıca erkeklerin beden duyumlarının kadınlardan farklı olabileceğini ifade etmişlerdir. Yapılan araştırmalarda, erkek öğrencilerin internet bağımlılığının ve internet kullanımının kız öğrencilere kıyasla manidar şekilde daha yüksek olduğu görülmektedir (Bayraktar, 2001; Kelleci vd., 2009; Kır & Sulak, 2014; Tutgun, 2009; Üçkardeş, 2010; Zorbaz, 2013). Ayrıca Akkoyunlu ve Orhan (2003) çalışmalarında üst düzey bilgisayar becerilerinin de erkekler lehine işlediğine işaret etmiştir. Erkeklerin daha yoğun bilgisayar kullanımının ve üst düzey beceriler bakımından da kızlara kıyasla daha yüksek düzeyde olmalarının bu durumu etkileyebileceği göz önünde bulundurularak Madde 6'da ortaya çıkan DMF'nin madde etkisinden kaynaklandığı; maddenin yanlılık içermediği yorumuna gidilmiştir.

“Madde 17. İnternet üzerinden yapılan sınavlar daha güvenilir sonuçlar verir.” ve “Madde 19. İnternet üzerinden yapılan sınavlarda hatalar daha az olur.” ifadelerini içeren maddeler doğrudan e-değerlendirme uygulamalarının güvenilirlik boyutuyla ilgili olup uzman görüşlerine göre bu maddelerdeki farklılığın temel sebebini az önce açıklandığı gibi teknoloji kullanım sıklığına ve internet bağımlılığına dayandırmak mümkündür. Bununla birlikte Ölçme ve Değerlendirme alanında doktorasını tamamlamış uzmanın önermiş olduğu, Başol vd. (2017) tarafından yapılan bir araştırmada, e-sınavların yaygınlaştırılması fikrine erkek öğrencilerin kız öğrencilere kıyasla daha olumlu yaklaştığı görülmüştür. Tüm bu faktörler göz önünde bulundurulduğunda, erkek öğrencilerin e-değerlendirmeye daha yüksek güvenilirlik atfetmesinin

sebebi de temellendirilmiş olmaktadır. Erkek öğrencilerin, interneti kız öğrencilere kıyasla daha sık kullandıkları, bu konuda daha yüksek bağımlılık puanlarına sahip oldukları, ayrıca e-değerlendirmelerle daha fazla karşılaşmak istedikleri ve bu sınavların güvenilirliklerinden kız öğrencilere kıyasla daha emin oldukları söylenebilir.

“Madde 13. İnternet üzerinden yapılan sınavda daha iyi konsantre olabilirim.” ve “Madde 14. İnternet üzerinden yapılan sınavlar benim kişilik özelliklerime daha uygundur.” ifadeleri temelinde uzman görüşleri ele alındığında, görsel zekâ, başarı odaklı motivasyon ve sınav kaygısı bakımından kız öğrencilerin ekran başında daha iyi konsantre olması ve buna dayanarak bu tür sınavları da kişilik özelliklerine daha uygun bulması beklenen bir sonuç olarak ortaya çıkmaktadır. Son olarak, uzmanların, “Madde 16. İnternet üzerinden yapılan sınav bütün becerilerimi ölçebilir.” ifadesi hakkında, PC sahibi olmayan grubun bu araca fazla anlam yüklediği veya bu aracı pek bilmediği şeklinde görüş bildirdiği görülmüştür. Uzman görüşleri sonucu bu maddede gözlenen DMF'nin de madde etkisinden kaynaklanabileceği düşünülmüştür. Alanyazında bu duruma ilişkin pek araştırmaya rastlanmamış olsa da Garland ve Noyes (2004, 2008) araştırmalarında bilgisayar deneyiminin ve kullanımının bilgisayara yönelik tutumun zayıf birer yordayıcısı olduğunu ortaya koyarken; Öz (2015) bilgisayar okuryazarlığının tutum üzerinde anlamlı bir etkisi olduğunu öne sürmüştür.

Bu araştırmanın bir sınırlılığı, e-değerlendirmeye yönelik tutum ölçeğine ilişkin cevapların çevrim içi ortamda toplanmış olmasıdır. Araştırmacılara yüz yüze şekilde veri toplayarak benzer çalışmaları tekrarlamaları önerilebilir. Ayrıca araştırmanın çalışma grubunu bir Anadolu üniversitesinde lisans düzeyinde öğrenim görmekte olan Eğitim Fakültesi öğrencileri oluşturmaktadır. Diğer araştırmaların büyük şehirlerde, farklı eğitim seviyelerinde ve farklı fakültelerde öğrenim gören öğrenciler üzerinden gerçekleştirilmesi; bu değişkenler temelinde madde yanlılığının incelenmesi önerilebilir. Bununla birlikte araştırmanın nitel kısmını oluşturan uzman grubundan görüşler de e-posta yoluyla edinilmiştir. Araştırmacılara yüz yüze bireysel görüşmeler, odak grup görüşmeleri veya Delphi tekniği ile uzman görüşlerinin edinilmesi önerilebilir.

Bu çalışmada yalnızca POLY-SIBTEST ve MantelTest sonucu yüksek düzeyde DMF içerdiği belirlenen maddeler yanlılık bakımından incelenmiş; başka bir yöntem ile DMF araştırılmamıştır. Bu çalışmadan elde edilen POLY-SIBTEST ve MantelTest bulguları bir arada ele alındığında, DMF analizinde işe koşulan bu iki yöntemin birbiriyle pek de uyumlu olmayan sonuçlar verdiği söylenebilir. Yöntemlerden elde edilen bulgulara göre, iki yöntem de Madde 6'nın erkek öğrenciler lehine yüksek düzeyde DMF içerdiğini ortaya koymuştur. Bununla birlikte uzmanların bir kısmının bu madde hakkında görüşlerini bir zemine oturtamadıkları gözlemlenmiştir. Bu bağlamda iki yöntem uyumsuz gibi görünse de yöntemlerin uzman görüşleriyle birleştikleri bir noktanın varlığından söz etmek mümkündür.

Bununla birlikte Madde 13 için POLY-SIBTEST kız öğrenciler lehine yüksek düzeyde DMF'ye işaret ederken MantelTest aynı madde için orta düzeyde DMF'nin varlığına bir kanıt sunmuştur. Alanyazın incelendiğinde kimi araştırmalarda DMF belirleme yöntemlerinin birbirleriyle oldukça uyumlu sonuçlar verdiği görülürken (Sırgancı, 2012; Şenferah, 2015); bazı araştırmalarda yöntemler arası uyumun düşük düzeyde sağlandığı veya sağlanmadığı dikkat çekmektedir (Atalay, 2010; Kuzu, 2021; Gök vd., 2010; Gür, 2019; Öğretmen, 2006; Yıldız, 2022). Bunun bir sebebi, yöntemlerin DMF düzeylerine ilişkin sınıflamada farklı kesme noktalarını dikkate almaları ve/veya DMF belirlemede farklı algoritmaları kullanmaları olabilir. Ayrıca özellikle POLY-SIBTEST yönteminin diğer yöntemlere kıyasla daha fazla sayıda DMF'li madde tespit ettiğini ve bu nedenle de daha duyarlı olduğunu belirten araştırmalar mevcuttur (Henderson, 2001; Kuzu, 2021; Sırgancı, 2012). Bu nedenle araştırmacılara ileride yapılacak çalışmalarda, farklı yöntemlerle DMF analizinin uygulanması ve sonuçların bir arada değerlendirilmesi önerilebilir. Bununla birlikte yapılacak çalışmalarda KTK'nın yanında, varsayımları sağlanarak MTK'ya dayalı yöntemlerin de ele alınması önerilebilir.

**Kaynakça**

- Akkoyunlu, B., & Orhan, F. (2003). Bilgisayar ve öğretim teknolojileri eğitimi (BÖTE) bölümü öğrencilerinin bilgisayar kullanma öz yeterlik inancı ile demografik özellikleri arasındaki ilişki. *The Turkish Online Journal of Educational Technology*, 2(3), 86-93.
- Arıkan Akın, Ç. (2015). Değişen madde fonksiyonu belirlemede MTK-olabilirlik oranı, ordinal lojistik regresyon ve POLY-SIBTEST yöntemlerinin karşılaştırılması. *E-Uluslararası Eğitim Araştırmaları Dergisi*, 6(1), 1- 16. <https://doi.org/10.19160/e-ijer.24504>
- Arıkan Akın, Ç., Uğurlu, S., & Atar, B. (2015). MIMIC, SIBTEST, lojistik regresyon ve Mantel Haenszel yöntemleriyle gerçekleştirilen DMF ve yanlılık çalışması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 31(1), 34-52. DOI:10.16986/HUJE.2015014226
- Asil, M., & Gelbal, S. (2012). PISA öğrenci anketinin kültürler arası eşdeğerliği. *Eğitim ve Bilim*, 37(166), 236-249.
- Atalay, K. (2010). *PISA 2006 öğrenci anketinde yer alan tutum maddelerinin değişen madde fonksiyonu açısından incelenmesi* [Yüksek lisans tezi]. Hacettepe Üniversitesi.
- Ayan, C. (2011). *PISA 2009 fen okuryazarlığı alt testinin değişen madde fonksiyonu açısından incelenmesi* [Yüksek lisans tezi]. Hacettepe Üniversitesi.
- Bahar, M. (2014). Üniversite öğrencileri için e-değerlendirme tutum ölçeğinin geliştirilmesi. *E-International Journal of Educational Research*, 5(2), 38-53.
- Bahar, M., & Asil, M. (2018). Attitude towards e-assessment: influence of gender, computer usage and level of education. *The Journal of Open, Distance and e-Learning*, 33(3), 221-237. <https://doi.org/10.1080/02680513.2018.1503529>
- Başokçu, T. O., & Öğretmen, T. (2013). Öğretmen öz yeterlilik ölçeğinde değişen madde fonksiyonlarının ağırlıklandırılmış cevap modeli ile belirlenmesi. *Ege Eğitim Dergisi*, 14(2), 63-78.
- Bilican Demir, S. (2014). *Çoklu puanlanan maddelerde madde işlev farklılığının Mantel Test ve olabilirlik oran testi ile karşılaştırılması* [Doktora tezi]. Ankara Üniversitesi.
- Bakan Kalaycıoğlu, D., & Kelecioğlu, H. (2011). Öğrenci seçme sınavının madde yanlılığı açısından incelenmesi. *Eğitim ve Bilim*, 36, 3-13.
- Başol, G., Kocadağ Ünver, T., & Çiğdem, H. (2017). Ölçme değerlendirme dersinde e-sınav uygulanmasına ilişkin öğrenci görüşleri. *Uluslararası Türk Eğitim Bilimleri Dergisi*, 2017(8), 111-128.
- Başaran, B., Yalman, M., & Gönen, S. (2016). Attitude scale towards web-based examination system (MOODLE) - Validity and reliability study. *Educational Research and Reviews*, 11(17), 1641-1649. DOI: 10.5897/ERR2016.2850
- Bayraktar, F. (2001). *İnternet kullanımının ergen gelişimindeki rolü* [Yüksek lisans tezi]. Ege Üniversitesi.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233-255.
- Çepni, Z. (2011). *Değişen madde fonksiyonlarının SIBTEST, Mantel-Haenszel, lojistik regresyon ve madde tepki kuramı yöntemleriyle incelenmesi* [Doktora tezi]. Hacettepe Üniversitesi.
- Yalçın, S. (2018). Karma yöntemler araştırmaları nasıl kullanılır? Ö. Çokluk Bökeoğlu (Ed.). *Karma yöntemler araştırması: Alana yönelik bir kılavuz içinde* (ss.105-133). Nobel Akademi Yayıncılık
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14, 464-504. <https://doi.org/10.1080/10705510701301834>
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative and mixed methods approaches* (4th ed.). SAGE Publications.
- Demir, S. (2013). *PISA 2009 matematik okuryazarlığı alt testinde bulunan maddelerinin Mantel Haenszel, SIBTEST ve lojistik regresyon yöntemleri ile değişen madde fonksiyonunun incelenmesi* [Yüksek lisans tezi]. Abant İzzet Baysal Üniversitesi.
- Doğan, N., & Öğretmen, T. (2008). Değişen madde fonksiyonunu belirlemede Mantel-Haenszel, ki-kare ve lojistik regresyon tekniklerinin karşılaştırılması. *Eğitim ve Bilim*, 33(148), 100-112.
- Erdem Keklik, D. (2012). *İki kategorili maddelerde tek biçimli değişen madde fonksiyonu belirleme tekniklerinin karşılaştırılması: Bir simülasyon çalışması* [Doktora tezi]. Ankara Üniversitesi.

- Garland, K. J., & Noyes, J. M. (2004). Computer experience: A poor predictor of computer attitudes. *Computers in Human Behavior*, 20, 823–840. <https://doi.org/10.1016/j.chb.2003.11.010>
- Garland, K. J., & Noyes, J. M. (2008). Computer attitude scales: How relevant today? *Computers in Human Behavior*, 24, 563–575. <https://doi.org/10.1016/j.chb.2007.02.005>
- Gierl, M. J., & Haladyna, T. M. (2013). *Automatic item generation: Theory & practice*. Routledge.
- Gök, B., Kelecioğlu, H., & Doğan, N. (2010). Değişen madde fonksiyonunu belirlemede Mantel Haenszel ve Lojistik Regresyon tekniklerinin karşılaştırılması. *Eğitim ve Bilim*, 35(156), 3-16.
- Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups?: Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care*, 44, 78-94. doi: 10.1097/01.mlr.0000245454.12228.8f.
- Guida, F., & Ludlow, L. (1989). A cross-cultural study of test anxiety. *Journal of Cross Cultural Psychology*, 20, 178-190.
- Gül, E., Çokluk, Ö., & Gül, Ç. D. (2015). Development of an attitudes scale toward online assessment. *Procedia – Social and Behavioral Sciences*, 174, 529–536. <https://doi.org/10.1016/j.sbspro.2015.01.699>
- Güler, D., & Çakır, G. (2013). Lise son sınıf öğrencilerinin sınav kaygısını yordayan değişkenlerin incelenmesi. *Türk Psikolojik Danışma ve Rehberlik Dergisi*, 4(39), 82-94.
- Gür, E. (2019). *PISA 2015 uygulamasındaki maddelerin kültüre göre değişen madde fonksiyonu açısından incelenmesi* [Yüksek lisans tezi]. Hacettepe Üniversitesi.
- Hambleton, R., & Rogers, J. (1994). Item bias review. *Practical Assessment, Research, and Evaluation*, 4, 1-3.
- Henderson, D. L. (1999). *Investigation of differential item functioning in exit Examinations across item format and subject area* [Doctoral dissertation]. University of Alberta.
- Henderson, D. L. (2001). *Prevalence of gender DIF in mixed format high school exit examinations*. Paper presented at the Annual Meeting of the American Educational Research Association (Seattle, WA, April 10-14, 2001). <https://eric.ed.gov/?id=ED458284>
- Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2022). “semTools: Useful tools for structural equation modeling.” R package version 0.5-6. Retrieved from <https://CRAN.R-project.org/package=semTools>
- Karakaya, İ., & Kutlu, Ö. (2012). Seviye Belirleme Sınavı'ndaki Türkçe alt testlerinin madde yanlılığının incelenmesi. *Eğitim ve Bilim*, 37(165), 342-368.
- Kavakçı, Ö., Güler, A. S., & Çetinkaya, S. (2011). Sınav kaygısı ve ilişkili psikiyatrik belirtiler. *Klinik Psikiyatri*, 14, 7-16.
- Kelleci, M., Güler, N., Sezer, H., & Gölbası, Z. (2009). Lise öğrencilerinde internet kullanma süresinin cinsiyet ve psikiyatrik belirtiler ile ilişkisi. *TAF Prev Med Bull*, 8(3), 223-230.
- Kır, İ., & Sulak, Ş. (2014). Eğitim fakültesi öğrencilerinin internet bağımlılık düzeylerinin incelenmesi. *Elektronik Sosyal Bilimler Dergisi*, 13(51), 150-167. <https://doi.org/10.17755/esosder.96255>
- Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). The Guilford Press.
- Korkmaz, M. (2005). *Madde cevap kuramına dayalı olarak çok kategorili maddelerde madde ve test yanlılığının (işlevsel farklılığın) incelenmesi* [Doktora tezi]. Ege Üniversitesi.
- Kuzu, Y. (2021). *Çok kategorili puanlanan maddelerde değişen madde ve adım fonksiyonu belirleme yöntemlerinin incelenmesi* [Doktora Tezi]. Hacettepe Üniversitesi.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741-749. <https://doi.org/10.1037/0003-066X.50.9.741>
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58, 525-543. <https://doi.org/10.1007/BF02294825>
- Meade, A. W., & Lautenschlager, G. J. (2004). A comparison of item response theory and confirmatory factor analytic methodologies for establishing measurement equivalence/invariance. *Organizational Research Methods*, 7(4), 361–388. <https://doi.org/10.1177/1094428104268>
- Meyer, J. P. (2014). *Applied measurement with jMetrik*. Routledge.



- McLean, C. P., & Anderson, E. R. (2009). Brave men and timid women? A review of the gender differences in fear and anxiety. *Clinical Psychology Review*, 29(6), 496-505. doi: 10.1016/j.cpr.2009.05.003.
- Öğretmen, T. (2006). *Uluslararası okuma becerilerinde gelişim projesi (PIRLS) 2001 testinin psikometrik özelliklerinin incelenmesi: Türkiye-Amerika Birleşik Devletleri örneği* [Yayımlanmamış doktora tezi]. Hacettepe Üniversitesi.
- Öz, H. (2015). An investigation of preservice English teachers' perceptions of mobile assisted language learning. *English Language Teaching*, 8(2), 22-34. <http://dx.doi.org/10.5539/elt.v8n2p22>
- R Development Core Team (2021). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing.
- Rosseel Y (2012). "lavaan: An R Package for Structural Equation Modeling." *Journal of Statistical Software*, 48(2), 1-36. [doi:10.18637/jss.v048.i02](https://doi.org/10.18637/jss.v048.i02).
- Roussos, L. A., & Stout, W. F. (1996). Simulation studies of the effects of small sample size and studied item parameters on SIBTEST and Mantel-Haenszel Type I error performance. *Journal of educational measurement*, 33(2), 215-230. <https://doi.org/10.1111/j.1745-3984.1996.tb00490.x>
- Sırgancı, G. (2012). *PISA 2006 öğrenci anketi madde yanlılığı sıralı lojistik regresyon ve POLY SIBTEST yöntemleri ile test edilmesi* [Yüksek lisans tezi]. Abant İzzet Baysal Üniversitesi.
- Shealy, R., & Stout, W. (1993). A model-based standardization approach that separates true bias/DIF from group ability differences and detects test bias/DTF as well as item bias/DIF. *Psychometrika*, 58(2), 159-194. <https://doi.org/10.1007/BF02294572>
- Şenferah, S. (2015). *2010 seviye belirleme sınavı matematik alt testi için değişen madde fonksiyonlarının ve madde yanlılığının incelenmesi* [Yüksek lisans tezi]. Gazi Üniversitesi.
- Tiryaki, F. (2019). *PISA 2015 öğrenci tutum anketlerinin değişen madde fonksiyonu ve ölçme değişmezliğinin incelenmesi* [Yüksek lisans tezi]. Ankara Üniversitesi.
- Tunç, E. (2016). *İki ve çok kategorili puanlanan maddelerde değişen madde fonksiyonlarının karşılaştırılması* [Doktora tezi]. Ankara Üniversitesi.
- Tutgun, A. (2009). *Öğretmen adaylarının problemleri internet kullanımlarının incelenmesi* [Yüksek Lisans Tezi]. Marmara Üniversitesi.
- Uzun Özer, B., & Topkaya, N. (2011). Akademik erteleme ve sınav kaygısı. *Adnan Menderes Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Dergisi*, 2(2), 12-19.
- Üçkardeş, E. A. (2010). *Mersin üniversitesi öğrencileri arasında internet bağımlılığının değerlendirilmesi* [Uzmanlık Tezi]. Mersin Üniversitesi.
- Ünsal Özberk, E. B., & Koç, N. (2017). WÇZÖ-IV maddelerinin cinsiyet ve sosyo-ekonomik düzey açısından işlev farklılığının belirlenmesinde kullanılan yöntemlerin karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 8(1), 112-127. <https://doi.org/10.21031/epod.287577>
- Yandı, A. (2017). *Ölçme eşdeğerliğini incelemede kullanılan yöntemlerin farklı koşullar altında istatistiksel güç oranları açısından karşılaştırılması* [Doktora tezi]. Ankara Üniversitesi.
- Yıldız, A. S. (2022). *Derecelendirme ölçeklerinde örneklem büyüklüklerine göre DMF yöntemlerinin farklı alt gruplarda karşılaştırılması* [Yüksek lisans tezi]. Akdeniz Üniversitesi.
- Yıldız, H. Y. (2007). *Sınav kaygısı-ana-baba tutumları ve mükemmeliyetçilik arasındaki ilişkinin incelenmesi* [Yüksek Lisans Tezi]. Gazi Üniversitesi.
- Zorbaz, O. (2013). *Lise öğrencilerinin problemleri internet kullanımının sosyal kaygı ve akran ilişkileri açısından incelenmesi* [Yüksek Lisans Tezi]. Hacettepe Üniversitesi.
- Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and likert-type (ordinal) item scores*. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.



## EKLER

### Ek-1. Cinsiyete ve PC'ye göre POLY-SIBTEST sonuçları

SIBTEST-Focal weighting									
Run no.	Suspect Item	Subtest Numbers	Beta estimate	standard error	p-value	p-elim		MS SSD	F L A G
1	1		0.113	0.093	0.226 E	.26	.46	0.47	0
2	2		-0.013	0.118	0.915 E	.21	.41	0.47	0
3	3		0.018	0.125	0.885 E	.17	.36	0.48	0
4	4		0.031	0.133	0.815 E	.23	.38	0.47	0
5	5		0.179	0.080	0.026 E	.25	.44	0.47	0
6	6		0.513	0.160	0.001 E	.14	.37	0.44	0
7	7		0.152	0.124	0.220 E	.19	.35	0.47	0
8	8		0.050	0.136	0.714 E	.19	.32	0.46	0
9	9		-0.168	0.111	0.132 E	.23	.42	0.48	0
10	10		-0.175	0.126	0.164 E	.17	.38	0.48	0
11	11		-0.128	0.120	0.288 E	.18	.35	0.48	0
12	12		-0.220	0.129	0.087 E	.24	.43	0.49	0
13	13		-0.275	0.129	0.032 E	.17	.36	0.49	0
14	14		-0.279	0.132	0.035 E	.14	.38	0.49	0
15	15		-0.003	0.106	0.979 E	.19	.41	0.48	0
16	16		-0.059	0.109	0.589 E	.17	.40	0.48	0
17	17		0.216	0.108	0.045 E	.15	.36	0.47	0
18	18		-0.115	0.104	0.268 E	.15	.34	0.48	0
19	19		0.282	0.132	0.032 E	.19	.36	0.46	0
20	20		-0.127	0.169	0.452 E	.23	.41	0.50	0
21	21		0.003	0.134	0.980 E	.19	.34	0.48	0
22	22		0.083	0.142	0.560 E	.20	.37	0.48	0
23	23		-0.128	0.097	0.187 E	.19	.31	0.48	0

SIBTEST-Focal weighting									
Run no.	Suspect Item	Subtest Numbers	Beta estimate	standard error	p-value	p-elim		MS SSD	F L A G
1	1		-0.038	0.068	0.580 E	.29	.15	0.18	0
2	2		-0.092	0.104	0.374 E	.25	.18	0.19	0
3	3		-0.008	0.101	0.940 E	.26	.17	0.18	0
4	4		-0.083	0.108	0.445 E	.25	.17	0.18	0
5	5		-0.078	0.064	0.226 E	.24	.15	0.18	0
6	6		0.102	0.103	0.321 E	.30	.15	0.17	0
7	7		0.109	0.117	0.348 E	.25	.18	0.17	0
8	8		0.021	0.096	0.826 E	.21	.11	0.18	0
9	9		0.143	0.093	0.122 E	.29	.13	0.17	0
10	10		-0.006	0.128	0.963 E	.31	.14	0.17	0
11	11		-0.004	0.092	0.966 E	.30	.13	0.17	0
12	12		0.087	0.103	0.399 E	.28	.14	0.17	0
13	13		0.071	0.096	0.461 E	.25	.19	0.17	0
14	14		0.005	0.088	0.955 E	.24	.11	0.17	0
15	15		0.059	0.093	0.528 E	.30	.15	0.16	0
16	16		-0.147	0.074	0.047 E	.26	.19	0.18	0
17	17		-0.132	0.087	0.132 E	.27	.13	0.19	0
18	18		0.167	0.094	0.075 E	.22	.13	0.16	0
19	19		-0.037	0.089	0.677 E	.27	.15	0.18	0
20	20		-0.171	0.140	0.220 E	.19	.13	0.19	0
21	21		0.200	0.106	0.059 E	.24	.13	0.15	0
22	22		-0.024	0.103	0.812 E	.26	.18	0.17	0
23	23		0.108	0.082	0.190 E	.22	.16	0.18	0

## Ek-2. Cinsiyete ve PC'ye göre MantelTest sonuçları

Item	Chi-square	p-value	Valid N	E.S. (95% C.I.)			Class
m1	5,18	0,02	421	-0,10	( -0,23,	0,03)	AA
m2	0,26	0,61	430	-0,13	( -0,32,	0,06)	AA
m3	0,04	0,83	419	0,04	( -0,13,	0,22)	AA
m4	0,31	0,58	425	-0,05	( -0,26,	0,16)	AA
m5	1,75	0,19	392	-0,09	( -0,22,	0,03)	AA
m6	23,78	0,00	430	-0,47	( -0,68,	-0,27)	CC-
m7	2,20	0,14	426	-0,05	( -0,24,	0,15)	AA
m8	4,11	0,04	431	-0,18	( -0,37,	-0,00)	AA
m9	0,54	0,46	419	0,06	( -0,09,	0,22)	AA
m10	0,31	0,58	424	0,06	( -0,16,	0,28)	AA
m11	0,21	0,64	428	0,05	( -0,12,	0,21)	AA
m12	6,62	0,01	420	0,22	( 0,02,	0,41)	BB+
m13	4,42	0,04	423	0,24	( 0,05,	0,42)	BB+
m14	4,35	0,04	429	0,16	( 0,00,	0,31)	AA
m15	0,28	0,60	424	0,06	( -0,11,	0,23)	AA
m16	0,06	0,81	421	0,00	( -0,14,	0,14)	AA
m17	0,72	0,40	423	-0,05	( -0,21,	0,10)	AA
m18	0,57	0,45	416	0,10	( -0,07,	0,28)	AA
m19	3,35	0,07	430	-0,15	( -0,31,	0,02)	AA
m20	6,74	0,01	435	0,24	( -0,03,	0,50)	BB+
m21	0,30	0,59	428	0,05	( -0,14,	0,25)	AA
m22	0,27	0,60	423	-0,06	( -0,25,	0,14)	AA
m23	0,89	0,34	419	0,06	( -0,11,	0,22)	AA

## Options

Matching Variable: tp  
DIF Group Variable: gender  
Focal Group Code: k  
Reference Group Code: e

Elapsed time: 0 secs, 50 msec

Item	Chi-square	p-value	Valid N	E.S. (95% C.I.)			Class
m1	1,34	0,25	428	-0,06	( -0,19,	0,06)	AA
m2	2,48	0,12	442	-0,12	( -0,29,	0,04)	AA
m3	0,37	0,54	441	-0,03	( -0,20,	0,14)	AA
m4	0,95	0,33	454	-0,14	( -0,33,	0,05)	AA
m5	0,77	0,38	390	-0,03	( -0,14,	0,09)	AA
m6	0,67	0,41	448	0,05	( -0,14,	0,24)	AA
m7	0,60	0,44	454	0,04	( -0,12,	0,21)	AA
m8	0,34	0,56	444	-0,02	( -0,20,	0,16)	AA
m9	2,34	0,13	446	0,07	( -0,06,	0,21)	AA
m10	0,00	0,97	444	-0,02	( -0,19,	0,15)	AA
m11	0,03	0,86	456	0,01	( -0,13,	0,16)	AA
m12	2,14	0,14	448	0,07	( -0,10,	0,25)	AA
m13	0,00	0,99	441	-0,04	( -0,22,	0,14)	AA
m14	0,28	0,60	446	-0,06	( -0,21,	0,09)	AA
m15	1,19	0,28	443	0,07	( -0,08,	0,23)	AA
m16	2,68	0,10	430	-0,09	( -0,24,	0,05)	AA
m17	2,46	0,12	430	-0,05	( -0,21,	0,11)	AA
m18	2,04	0,15	424	0,15	( -0,01,	0,31)	AA
m19	0,02	0,90	442	0,09	( -0,07,	0,26)	AA
m20	0,92	0,34	453	-0,06	( -0,31,	0,18)	AA
m21	3,28	0,07	449	0,13	( -0,05,	0,31)	AA
m22	0,48	0,49	448	0,03	( -0,14,	0,20)	AA
m23	0,04	0,84	431	-0,01	( -0,15,	0,14)	AA

## Options

Matching Variable: tpc  
DIF Group Variable: pcpc  
Focal Group Code: v  
Reference Group Code: y

Elapsed time: 0 secs, 48 msec

## EXTENDED ABSTRACT

### 1. INTRODUCTION

Due to the Covid-19 pandemic, the education process took place in an online environment both in Turkey and all around the world. Moreover, measurement and evaluation practices in education were similarly moved to the online environment and the concept of electronic assessment (e-assessment) gained importance in this period. Considering the needs of the age, it can be predicted that e-assessment applications will become more widespread in the upcoming years. In this regard, it can be suggested that analyzing the attitudes of students and even future teachers toward e-assessment based on different variables is among the issues that will continue to be important in the present and in future. Information on individuals' psychological characteristics such as attitude, perception, and interest are obtained indirectly through scales. Scales are expected to provide valid and highly reliable measurements. At the stage of collecting evidence for the validity of the measurement tool, it can be considered a necessity to examine the measurement tool in terms of invariance and bias. Cheung and Rensvold (2002) stated that measurement invariance is critical when comparing groups. If measurement invariance cannot be achieved, it will not be certain whether the differences between the groups are due to actual differences or different psychometric responses to scale items; thus, the results will not be interpreted (Cheung & Rensvold, 2002). In particular, it can be suggested that item bias will emerge if scalar invariance is not achieved. Item bias is defined as individuals of the same ability with different genders and cultural or ethnic backgrounds performing at different levels on an item (Hambleton & Rodgers, 1994). Therefore, it is a systematic error and reduces the validity. The differential item functioning (DIF) analyses are used in the detection of item bias. This study aims to analyze the e-Assessment Attitude Scale (EAAT), which was developed by Bahar (2014) and revised by Bahar & Asil (2018), in terms of measurement invariance, DIF, and item bias based on gender and possession of a personal computer.

### 2. METHOD

Since this study used both quantitative and qualitative methods, it can be said that the research has a mixed type (Creswell, 2014). Based on the fact that the quantitative and qualitative stages were applied sequentially and deductions were made by integrating the two result sets, it can be suggested that the research was conducted in a sequential quantitative -> qualitative design (Yalçın, 2018). The study group in the quantitative dimension of the study consisted of 537 pre-service teachers studying at the Faculty of Education of a state university in the 2021-2022 academic year. In the qualitative part of the study, expert opinions were consulted to determine whether the items that were found to contain high levels of DIF showed bias or not. The expert group consisted of faculty members working in various departments of the Faculty of Education and researchers who completed their doctorates. In this study, the e-Assessment Attitude Scale (EAAS) was used as the data collection tool. The 5-point Likert scale included 23 items in the dimensions of Type of Exam, Individual/Physical Characteristics, Technical Factors, and Social/Environmental Factors. The Cronbach's alpha coefficient of the EAAS was 0.85 (Bahar, 2014). A demographic information form prepared by the researcher was also applied to the study group together with the EAAS. During the qualitative part of the study, expert opinions were collected with a form consisting of open-ended questions on the scale items that were found to demonstrate high levels of DIF. Before proceeding to measurement invariance and DIF analyses, the reliability of the obtained data set was calculated. The McDonald's Omega coefficient for the entire scale was 0.95. Afterward, the assumptions were tested and the structure was verified using Confirmatory Factor Analysis (CFA) in each subgroup of the independent variables of the study, and it was observed that the model data fit was achieved in the subgroups. In the first stage, measurement invariance was tested in terms of gender and possession of a personal computer (PC) variables by applying multi-group CFA in the R (2021) environment. In the second stage, POLY-SIBTEST (Shealy & Stout, 1993) and MantelTest (Meyer, 2014) were used to identify the items

demonstrating DIF. Lastly, in the qualitative part of the study, expert opinions were obtained for the items demonstrating high levels of DIF according to both POLY-SIBTEST and MantelTest.

### 3. FINDINGS, DISCUSSION AND RESULTS

It was observed that EAAS did not provide metric invariance based on gender and PC variables. The POLY-SIBTEST results indicated that the items 5, 6, 17, 19 and the items 13, 14 demonstrated a high level of DIF in favor of male and female students, respectively. It was further determined that item 16 demonstrated high levels of DIF in favor of students who did not have a PC. The MantelTest showed that the items 12, 13 and 20 demonstrated a moderate level of DIF, and Item 6 demonstrated a high level of DIF in favor of male students. According to the PC variable, it was determined that all items showed a negligible level of DIF. Upon examining the opinions of the experts on the items that were determined to contain high levels of DIF according to both methods, it was seen that the experts agreed on the item effect point for all the items, except for one item (Item 6). It can be said that the opinions of the experts differ for item 6 (*My eyes get more tired during the online exam*). While one of the experts stated that the item should be questioned, it was determined that five experts thought this was an expected result, and two remained neutral. The experts who expressed similar opinions stated that computer use is more common among males, and thus, males would feel more comfortable in e-assessment applications, and those body sensations of males may differ from that of females. Relevant studies have shown that male students' internet addiction and internet use are significantly higher than those of female students (Bayraktar, 2001; Kelleci et al., 2009; Kır & Sulak, 2014; Tutgun, 2009; Üçkardeş, 2010; Zorbaz, 2013). Additionally, Akkoyunlu and Orhan (2003) pointed out that high-level computer skills are also in favor of males. Accompanied by expert opinions and the relevant literature, it was stated that DIF emerging in Item 6 was caused by the item effect, and it was interpreted that the item did not show bias.

## ARAŞTIRMANIN ETİK İZİNİ

Bu çalışmada “Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesi” kapsamında uyulması gerektiği belirtilen tüm kurallara uyulmuştur. Yönergenin ikinci bölümü olan “Bilimsel Araştırma ve Yayın Etiğine Aykırı Eylemler” başlığı altında belirtilen eylemlerden hiçbiri gerçekleştirilmemiştir.

### Etik kurul izin bilgileri

Etik değerlendirmeyi yapan kurul adı: Aksaray Üniversitesi İnsan Araştırmaları Etik Kurulu

Etik değerlendirme kararının tarihi: 21.12.2021

Etik değerlendirme belgesi sayı numarası: 2021/08-50

## ARAŞTIRMACILARIN KATKI ORANI

Bu araştırmanın tümü tek yazar tarafından tasarlanmış, yürütülmüş ve raporlanmıştır.

## ÇATIŞMA BEYANI

Bu çalışmada çıkar çatışması bulunmamaktadır.