



Farklı Ortamlarda Uygulanan Likert Tipi Ölçek ile Metrik Ölçeğin Psikometrik Özelliklerinin İncelenmesi*

Ufuk AKBAŞ**, Şeyhmus AYDOĞDU***, Şener BÜYÜKÖZTÜRK****

Makale Bilgisi	ÖZET
Geliş Tarihi: 13.08.2018	<p>Bu araştırmanın amacı kâğıt - kalem ve internet ortamlarında uygulanan iki farklı ölçeğin (Araştırmanın Yöntem Bölümüne İlişkin Öz Yeterlik Ölçeği - AYBİÖYÖ ve Araştırmaya Yönelik Kaygı Ölçeği - AYKÖ) psikometrik özelliklerinin incelenmesidir. Bu amaçla, AYBİÖYÖ ve AYKÖ'nün beş dereceli Likert tipi ölçek formu ile 10 cm'lik metrik formu önce kâğıt - kalem ortamında, daha sonra ise araştırma kapsamında hazırlanan bir sayfa üzerinden internet ortamında 391 lisans öğrencisine uygulanmıştır. Lisansüstü öğrencilerde kullanılmak üzere geliştirilmiş olan ve iki faktörden oluşan AYBİÖYÖ'nün iki faktörlü yapısının, lisans düzeyinde de bir maddenin diğer faktörde yer alması şeklindeki farklılık dışında aynen ortaya çıktığı; bir faktörden oluştuğu bilinen AYKÖ'nün ise bu yapıyı koruduğu ve ölçeklerin farklı veri setleri üzerinden hesaplanan güvenilirliklerinin yüksek olduğu görülmüştür. Her iki ölçek için, kâğıt - kalem ortamında uygulanan Likert tipi ölçekle elde edilen açıklanan varyans oranlarının ve güvenilirlik katsayılarının metrik ölçeğe ve internet ortamında gerçekleştirilen uygulamalara göre daha düşük olduğu belirlenmiştir. İki ucu etiketlenmiş metrik ölçeğin Likert tipi ölçek yerine kullanılabilmesi; kategori sayısına göre yapılan incelemelerde ise ideal kategori sayısının yedi olduğu görülmüştür.</p> <p>Anahtar Sözcükler: Metrik ölçek, likert tipi ölçek, araştırma, öz yeterlik, kaygı</p>
Kabul Tarihi: 04.03.2019	
Erken Görünüm Tarihi: 06.03.2019	
Basım Tarihi: 31.01.2020	

Investigation of Psychometric Traits of Metric Scale and Likert Type Scale Applied in Different Conditions

Article Information	ABSTRACT
Received: 13.08.2018	<p>The purpose of this study is to examine the psychometric properties of two different scales (Self-Efficacy Scale about the Method Sections of the Research Papers - SESMSR and The Research Anxiety Scale - TRAS) applied on paper - pencil tests and web. For this purpose, SESMSR and TRAS were applied to 391 undergraduate students in a five - point Likert type scale form and a 10 cm metric form in a paper - pencil environment and then on a web page prepared within the context of the research. The two-factor structure of SESMSR, which is developed to use for the graduate students, appeared exactly the same except for the fact that one item is included in the other factor at the undergraduate level; and it was seen that TRAS, which is known for consisting of one factor, maintained this structure and that the reliability of scales calculated over different data sets is high. For both scales, it was determined that the extracted variance and reliability coefficients obtained with the use of Likert type scale applied as paper - pencil test were lower than the metric scale and the web-based applications. It has been seen that the metric scale which is labeled on both ends can be used instead of the Likert-type scale and the ideal number of category is seven according to investigations.</p> <p>Keywords: Metric scale, likert scale, research, self-efficacy, anxiety</p>
Accepted: 04.03.2019	
Online First: 06.03.2019	
Published: 31.01.2020	

doi: 10.16986/HUJE.2019050088

Makale Türü (Article Type): Araştırma Makalesi

Kaynakça Gösterimi: Akbaş, U., Aydoğdu, Ş., & Büyüköztürk, Ş. (2020). Farklı ortamlarda uygulanan likert tipi ölçek ile metrik ölçeğin psikometrik özelliklerinin incelenmesi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 35(1), 222-242. doi: 10.16986/HUJE.2019050088

Citation Information: Akbaş, U., Aydoğdu, Ş., & Büyüköztürk, Ş. (2020). Investigation of psychometric traits of metric scale and likert type scale applied in different conditions. *Hacettepe University Journal of Education*, 35(1), 222-242. doi: 10.16986/HUJE.2019050088

* Bu çalışma 6. Uluslararası Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi'nde sözlü bildiri olarak sunulmuştur.

** Dr. Öğr. Üyesi, Hasan Kalyoncu Üniversitesi, Eğitim Fakültesi, Temel Eğitim Bölümü, Sınıf Öğretmenliği Programı, Gaziantep-TÜRKİYE. e-posta: ufuk.akbas@hku.edu.tr (ORCID: 0000-0002-6122-154X)

*** Dr. Öğr. Üyesi, Nevşehir Hacı Bektaş Veli Üniversitesi, Eğitim Fakültesi, Bilgisayar ve Öğretim Teknolojileri Eğitimi Bölümü, Bilgisayar ve Öğretim Teknolojileri Eğitimi A.B.D., Nevşehir-TÜRKİYE. e-posta: aydogduseyhmus@gmail.com (ORCID: 0000-0002-9075-8055)

**** Prof. Dr., Hasan Kalyoncu Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Rehberlik ve Psikolojik Danışmanlık Programı, Gaziantep-TÜRKİYE. e-posta: sener.buyukozturk@hku.edu.tr (ORCID: 0000-0002-0898-1697)

1. GİRİŞ

Psikolojik değişkenlerin ölçülmesinde farklı cevaplama ve puanlama şekline sahip ölçme araçlarından yararlanılmaktadır. Likert ölçeği tutum ölçekleri bu araçların içinde en yaygın olarak kullanılanıdır. Bunun nedeni, Likert tipi ölçeklerin geliştirilmesinin diğer ölçeklere göre daha kolay ve kullanışlılığının yüksek olmasıdır (Tavşancıl, 2002). Likert tipi ölçek geliştirme, tutum ölçmek amacıyla geliştirilmiş bir yöntem olmakla birlikte, başka psikolojik değişkenlerin ölçeklenmesi için de elverişli bir yöntemdir (Erkuş, 2016).

Avantajlarının ve yaygın kullanımının yanında Likert tipi ölçeklerde, elden edilen verilerin sıralama ölçeğinde olmasına rağmen eşit aralıklı kabul edilmesi gibi tartışmalı bir durum söz konusudur. Bu durum, özellikle parametrik istatistiksel teknikler için söz konusu olan ve değişkenin eşit aralıklı ve sürekli olması yönündeki varsayımların karşılanmasını güçleştirmektedir. Ayrıca, Likert tipi ölçeklerde dereceler arasında orta noktanın bulunup bulunmayacağı, kategorilerin nasıl adlandırılacağı, tüm kategorilerin adlandırılıp adlandırılmayacağı ve kategori sayısının kaç olacağı gibi hususlarda bazı kararların alınması gerekmektedir. Literatür incelendiğinde burada belirtilen konulardaki tartışmaların uzun zamandır süregeldiği görülmektedir.

Sosyal bilimlerde psikoloji ve eğitim araştırmalarında Likert tipi derecelenmeli ölçekler kullanılarak ölçülen tutum, kaygı, ilgi vb. değişkenlerin, gerçekte sıralama ölçeğine girmekle birlikte, araştırmacılar tarafından daha güçlü istatistikler kullanabilmek amacıyla aralık ölçeğinde kabul edildiği görülmektedir (Büyüköztürk, 2018). Bu istatistikler uygulandığında sıralama ölçeğindeki verilerin ardışık aralıklar arasındaki uzunluklarının eşit olduğu varsayılmaktadır. Bu durumda elde edilen istatistikler, ardışık aralıkların uzunluklarının farklılığı ölçüsünde hata taşırlar (Baykul, 2000).

Literatürde Likert tipi ölçeklerde bulunması gereken kategori sayısı ile ilgili tartışmalar bulunmaktadır. Harpe (2015) en az beş kategori içeren maddelerin ve ölçekten alınan toplam puanların sürekli kabul edilebileceğini; Tabachnick ve Fidell (2013) ise kategori sayısının yedi veya daha fazla olduğu ve buna ek olarak elde edilen verilerin istatistiksel analizlerin diğer varsayımlarını karşılaması halinde sürekli değişken olarak analizlere dâhil edilebileceğini belirtmektedir.

Nunnally ve Bernstein (1974) gerçekte matematiksel olarak sürekli olmasa bile 11 ya da daha fazla düzeyi olan değişkenlerin sürekli kabul edilebileceğini ve kategori sayısının azaltılmasına bağlı olarak kaybedilecek bilginin, cevaplayıcıların bu düzeyleri ne derece ayırt edebildiklerine bağlı olduğunu ifade etmektedir. Erkuş (2016) kategori sayısı azaldıkça ölçmenin duyarlılığının azalacağını arttıkça da ölçmenin duyarlılığının artacağını fakat sayının çok fazla olduğu durumlarda kategoriler arası ayırt etme algısının azalacağını ifade etmektedir.

Bununla beraber DiStefano (2002), maddelerin beş dereceli olduğu durumda normal dağılım varsayımı karşılanırsa bile negatif yanlı parametre kestirimlerinin elde edildiğini; dağılım normalden uzaklaştıkça da bu yanlılığın arttığını belirlemiştir. Sıklıkla rapor edilen Cronbach α katsayısının Likert tipi ölçeklerde kullanımı eleştirilmekte ve polikorik korelasyon matrisine dayalı iç tutarlılık katsayıları önerilmektedir (Zumbo, Gadermann ve Zeisser, 2007). Pedhazur ve Schmelkin (1991) ise başarı, zihinsel yetenek, tutum vb. özelliklerin sıralama ve eşit aralıklı ölçek arasında bir düzeyde ölçüldüğünü ve sıralama düzeyinde kabul edilerek işlem yapılması halinde ulaşılan bilgide önemli kayıpların yaşanacağını belirtmektedir. Norman (2010) da buna benzer şekilde, parametrik tekniklerin güçlü (robust) olduklarını ve bu teknikler ile Likert tipi verilerin analiz edilebileceğini savunmaktadır.

Dereceli ölçeklerde orta noktanın bulunup bulunmayacağı, kategorilerin nasıl adlandırılacağı, tüm kategorilerin adlandırılıp adlandırılmayacağı ve derece sayısının kaç olacağı gibi konulardaki tartışmaların oldukça eskiye dayandığı görülmektedir. İlgili literatür incelendiğinde ölçeklerde orta noktaya yer verilip verilmemesi konusunda bir fikir birliğinin bulunmadığı görülmektedir. Churchill ve Peter'e (1984) göre, ölçülen yapıya ilişkin tarafsız bir duygu ya da düşünceye sahip cevaplayıcıların bu durumlarını yansıtılmalarına fırsat verilmesi ve olumlu/olumsuz bir taraf seçmek zorunda bırakılmaları gerekmektedir. Orta noktaya yer verilmeyen uygulamalarda cevaplayıcılar taraf seçmede zorlanabilir ve bu durum, ölçme hatasını artırarak güvenilirliği düşürebilir.

Armstrong (1987) beş dereceli bir ölçekte üç puana karşılık gelen orta noktanın sıklıkla "kararsızım" şeklinde adlandırıldığını fakat "kararsızlık" (undecided) ile "tarafsızlık" (neutral) ifadelerinin aynı kavramlara karşılık gelmediğini ifade etmektedir. Tarafsızlık, katılma ile katılmama arasındaki orta noktaya; kararsızlık ise cevaplayıcının bilgi eksikliğine veya ilgisizliğine bağlı bir belirsizliğe karşılık gelmektedir. Benzer şekilde Raaijmakers, Hoof, Hart, Verbogt ve Vollenberg (2000) ve Baka, Figgou ve Triga (2012) tarafından yürütülen çalışmalar orta noktanın, konuyla ilgili bilgisinde eksikler olan kişiler tarafından seçildiğini ortaya koymaktadır.

Bardakçı, Haşiloğlu ve Balce (2015) kendisine sunulan bir konuda bilgisi olmayan cevaplayıcının "bihaber cevaplayıcı", verilen konu hakkında net bir fikir bildirmek istemeyen ya da tarafsız olmak isteyen cevaplayıcının ise "bitaraf cevaplayıcı" olduğunu ve zorlanmış (forced-choice - orta noktası bulunmayan) ölçek kullanmak için geçerli argümanlar bulunmadığı sürece, zorlanmamış (orta noktası bulunan) ölçeğin kullanımının daha uygun olduğunu ifade etmektedir.

Coelho ve Esteves (2007) ise bir grup ölçeğin beş ve on derece içeren formlarını karşılaştırdıkları araştırmada, beş dereceli ölçeğin orta noktasının on dereceli ölçeğin orta noktalarına (beş ve altı) göre daha fazla tercih edildiğini ve bunun, cevaplayıcıların çaba harcamak (response effort) istemedikleri için ortayı seçmeleriyle ilgili olduğunu ifade etmekte; on dereceli formun beş dereceliye göre daha üstün psikometrik özelliklere sahip olduğunu ve orta nokta kullanımının zorunlu olmaması gerektiğini savunmaktadırlar. Bora Semiz ve Altunışık (2016) ve Mattel ve Jacoby (1972) tarafından yapılan araştırmalarda da bu duruma benzer şekilde, kategori sayısı arttıkça orta noktaya yönelen tepkilerin azaldığı belirlenmiştir.

Erkuş, Sanlı, Güven ve Bağlı (2000) tarafından yürütülen Öğretmenliğe İlişkin Tutum Ölçeği'ni geliştirme çalışmasında, aynı maddeler beş dereceli form ile ve "kararsızım" seçeneğinin kaldırıldığı dört dereceli form ile iki farklı gruba uygulanmıştır. Her iki form için madde ayırt ediciliklerinin kabul edilebilir düzeyde ve dört dereceli form için genel olarak daha yüksek olduğu; faktör analizi sonucunda nihai ölçekte kalmasına karar verilen maddelerin de büyük ölçüde örtüştüğü görülmüştür. Nihai ölçeğin Cronbach alfa katsayılarının da oldukça yüksek ve birbirine yakın (dört derece için .96, beş derece için .99) olduğu görülmüş ve ölçeğin dört ya da beş dereceli olmasının geçerlik ve güvenilirlik açısından önemli bir fark yaratmadığı sonucuna ulaşılmıştır.

Churchill ve Peter (1984) tarafından yürütülen bir meta-analizde de orta noktanın bulunup bulunmamasının güvenilirlik üzerinde manidar bir etkisinin olmadığı görülmüştür. Bazı araştırmacılar da zorlanmış ya da zorlanmamış ölçeğin kullanımının araştırmacının inisiyatifinde olduğunu belirtmektedir (Hughes, 1969; Garland, 1991).

Ölçek kategorilerinin etiketlenmesi, ölçülen özelliğe ve maddelerin bağlamına göre değişebilmektedir. Diğer bir deyişle uygun bulma, önemli bulma, katılma düzeyi veya sıklığa yönelik ölçüm elde etmek amacıyla hazırlanmış ölçeklerde farklı kategori etiketlerinin bulunması doğal ve beklenen bir durumdur. Tepkilerin genel olarak düşükten yükseğe sıralandığı tasarımlar kullanılmakla beraber, ilk kategorinin en düşük tepki düzeyine karşılık geldiği örneklere de rastlanmaktadır. Türkçe ölçekler için kategorilerin nasıl etiketlenebileceğine ilişkin öneriler için çeşitli kaynaklardan yararlanılabilir (Bayat, 2014; Erkuş, 2016; Tavşancıl, 2002; Tezbaşaran, 1996; Turan, Şimşek ve Aslan, 2015).

Erkuş (2016) farklı dereceleme türlerinden hangisinin kullanılacağına amaca, ölçülen özelliğe ve madde köküne bağlı olduğunu belirtmektedir. İfade ile tepki kategorinin etiketleri birbirine uyumlu ve mümkünse tüm maddeler için kategori etiketlerinin aynı olması gerekir. Özellikle davranışsal sıklık belirten kategori etiketlerinde, dörtten fazla kategorinin özellikle Türkçede sorun çıkarmaktadır. Örneğin "ara sıra", "bazen", "nadiren", "seyrek olarak" ve "genellikle", "çoğunlukla", "sık sık" gibi zaman niteleyicileri arasındaki fark ayırt edilememektedir.

Derecelerin sadece sayısal ifadelerden oluştuğu örneklere de rastlanmaktadır. Uyumaz (2013) tarafından yapılan araştırmada, kategorilerin sözel ya da sayısal ifadelerle etiketlenmesinin güvenilirlik ve madde ayırt ediciliği üzerinde genel olarak fark yaratmadığı; doğrulayıcı faktör analizi ile yapılan incelemelerde elde edilen bazı uyum indekslerinin sözel ifadeler lehine daha yüksek olduğu tespit edilmiştir. Araştırmacıların ölçekteki dereceleri sözel olarak açık bir şekilde ifade etmesi, cevaplayıcıları, sayılara yüklenen anlamın karıştırıcı etkisinden kurtaracak ve ölçeğin psikometrik niteliklerini arttıracaktır.

Churchill ve Peter (1984) bütün kategorilerin etiketlenmesinin cevaplayıcının yaşayabileceği karışıklıkları önleyeceğini, belirsizlikleri ortadan kaldıracığını ve buna bağlı olarak da tüm kategorilerin etiketlendiği durumlarda daha yüksek güvenilirlik değerlerinin elde edilmesinin beklenebileceğini ifade etmektedir. Öte yandan, 1964-1982 yılları arasında yayınlanan 6484 makale üzerinden yürüttükleri meta-analiz çalışmasında durumun böyle olmadığını, sadece üç noktaların etiketlendiği ölçeklerin tüm kategorilerin etiketlendiği ölçeklere göre daha yüksek güvenilirliğe sahip olduğunu görmüşlerdir.

Finn (1972) birinci formda tüm kategorileri, ikinci formda sadece ilk ve son kategorileri, üçüncü formda ise sadece ikinci ve dördüncü kategorileri etiketlenmiş olan beş dereceli bir ölçekten elde edilen verileri karşılaştırmıştır. Formlardan elde edilen aritmetik ortalamalar ve güvenilirlik katsayıları arasında manidar bir farkın olmadığı görülmüş ve etiketlenmenin, dereceli ölçeklerin tasarlanması sürecinde önemli bir husus olmadığını belirtmiştir. Benzer şekilde Dixon, Bobo ve Stevick (1984) tüm kategorilerin etiketlenmesi ile sadece üç kategorilerin etiketlenmesi arasında fark olmadığını belirtmektedir.

Weng (2004) kategori sayısı 3 - 9 arasında değişen ve tüm kategorilerin etiketlendiği ve sadece ilk ve son kategorilerin etiketlendiği iki farklı ölçeğin iç tutarlılık ve test - tekrar test güvenilirliklerini incelediği araştırmada, genel olarak kategori sayısına düştükçe iç katsayılarının ve tekrar test güvenilirliklerinin de düştüğünü belirlemiştir. Bununla birlikte, araştırmada kullanılan iki farklı ölçekten elde edilen verilerin güvenilirliğindeki değişimin (ya da değişmemenin) asıl belirleyicisinin, ölçek maddelerinin faktör yükleri olduğunu belirtmektedir. Daha yüksek faktör yüküne sahip maddelerden oluşan ölçeğin güvenilirliği, kategori sayısından ve kategorilerin etiketlenmemesinden daha az etkilenmektedir. Bir başka ifadeyle, kategorilerin nasıl etiketleneceği önemli bir husus değildir. Cummins ve Gullone (2000) ise tüm kategorilerin etiketlenmesinin dereceli ölçeklerin halihazırda eleştirilen eşit aralıklı olduğu şeklindeki kabulü gereksiz bir şekilde zayıflattığını öne sürmekte ve sadece üç noktaların etiketlenmesini önermektedir.

Likert tipi ölçeklerde kaç kategoriye yer verilmesi gerektiğine yönelik tartışmaların ise bizzat Rensis Likert'e kadar dayandığı ve sunulan önerilerin çok da tutarlı olmadığı görülmektedir. Murphy ve Likert (1938) güvenilirliğin, madde veya derece

sayısına bağlı olarak artış göstereceğini belirtmektedir. Üç dereceli 44 maddeden elde edilen korelasyon değeri .88 iken, beş dereceli 36 maddeden elde edilen değer .94'tür (akt. Cox III, 1980).

Jacoby ve Matell (1971) kategori sayısının 2 – 19 arasında manipüle ettikleri araştırmalarında, iç tutarlılık ve test – tekrar test güvenilirliği ile ölçüt bağıntılı geçerlik temelinde incelemeler gerçekleştirmişlerdir. Kategori sayısının güvenilirlik ve geçerlik üzerinde tutarlı bir etkisi olmadığına dayanarak iki ya da üç kategorinin yeterli olduğunu ifade etmişlerdir.

Finn (1972) üç, beş, yedi ve dokuz dereceli formları karşılaştırmış ve en yüksek güvenilirlik değerlerini beş ve yedi, en düşük değerleri ise üç ve dokuz derece için elde etmiştir. McMordie (1979) özgün formu ikili puanlanan Ölüm Kaygısı Ölçeği'ni yedi dereceli form ile karşılaştırmış ve yedi dereceli formun güvenilirliğinin daha yüksek olduğunu belirlemiştir. Cicchetti, Showalter ve Tyrer (1985) tarafından yapay veriler üzerinden yürütülen çalışmada kategori sayısı 2 – 100 arasında manipüle edilmiş ve güvenilirliğin kategori sayısına bağlı olarak düzenli bir şekilde arttığı belirlenmiştir. En hızlı artış iki kategoriden üç kategoriye geçildiğinde yaşanırken, yedinci kategoriden itibaren önemli bir farkın olmadığı belirlenmiştir.

Bandalos ve Enders (1996) dağılım özelliklerini ve maddeler arası korelasyon değerlerini manipüle ederek yürüttükleri çalışmada üç, beş, yedi, dokuz ve 11 kategoriden oluşan veriler üzerinden Cronbach alfa katsayılarını karşılaştırmışlardır. Maddeler arasındaki korelasyonlarla birlikte güvenilirlik katsayılarının az da olsa artış gösterdiği ve bu durumun puanların dağılımına ve kategori sayısına bağlı olmadığı belirlenmiştir. Diğer bir deyişle, güvenilirliği artıran kategori sayısı değil ölçeğin iç yapısının niteliğidir.

Dawes (2008) beş, yedi ve on dereceli formlardan elde edilen verileri betimsel istatistikler ve dağılım özellikleri üzerinden karşılaştırmıştır. Beş ve yedi dereceli formlar arasında manidar bir farklılık yok iken on dereceli formun ortalamasının daha düşük olduğu; çarpıklık ve basıklık değerlerinin ise farklı formlar için benzer olduğu görülmüştür. Leung (2011) dört, beş, altı ve 11 dereceli formlarla elde edilen verilerde betimsel istatistik, ayırt edicilik, güvenilirlik ve faktör yükleri açısından fark olmadığını, normallığe en yakın sonuçların 11 derece ile elde edildiğini belirtmektedir.

Kategori sayısının 21 olduğu durumlarda iyi sonuçlar elde edilebileceğini belirten araştırmalara (Pearse, 2011) da rastlanmakla beraber Peterson (1994) ve İnal, Yılmaz Koğar, Demirdüzen ve Gelbal (2017) tarafından yürütülen meta-analizlerde, en sık karşılaşılan kategori sayısının beş olduğu görülmüştür. Lissitz ve Green (1975), Lozano, Garcia-Cueto ve Muniz (2008), Preston ve Colman (2000) ve Wakita, Ueshima ve Noguchi (2012) tarafından yürütülen çalışmalar, yedi kategorinin en iyi sonuçları verdiğini ortaya koymaktadır. Cox III (1980) farklı ölçekler için tüm koşullarda kullanılacak tek bir kategori sayısının önerilemeyeceğini fakat bunun yanında 7±2 kategorinin “sihirli” olduğunu ifade etmektedir. Burada bahsedilen çalışmalar bir bütün olarak değerlendirildiğinde de bu ifadenin desteklendiği anlaşılmaktadır.

Bu noktada görsel analog ölçek veya grafik derecelendirme ölçeği gibi adlarla anılan metrik ölçekler, Likert tipi ölçeklere önemli bir alternatif olarak karşımıza çıkmaktadır. Farklı türleri olmakla birlikte tipik olarak ilgili maddeyi okuyan cevaplayıcının bir doğru parçasının iki ucu arasında bir noktayı işaretlemesi ve bu noktanın, doğru parçasının başlangıcına olan uzaklığının ölçülmesi şeklinde uygulanmaktadır. Kerlinger (1986) metrik ölçeklerin, dereceli ölçekler içinde en iyisi olduğunu belirtmektedir. Cevaplayıcının veya gözlemcinin zihninde süreklilik sağlamak, eşit aralıklar sunmakta, anlaşılabilirlik ve kullanılabilirlik bakımından kolay ve sade bir yapı sergilemekte (akt. Cook, Heath, Russel, Thompson ve Thompson, 2001) ve orta noktanın bulunup bulunmayacağı yönündeki tartışmayı sonlandırmaktadır (Reips ve Funke 2008). Metrik ölçek ile elde edilen verilerin oran ölçeğinde sayılabileceğini öne süren araştırmalara da rastlanmaktadır (Myles, Troedel, Boquest, Reeves, 1999; Price, McGrath, Ruff ve Buckingham, 1983). Farklı tasarımlarla kullanılacak metrik ölçek türleri ile ilgili açıklamalar Tablo 1'de verilmiştir (Atan ve Yavuz, 2010).

Tablo 1.

Metrik ölçek türleri ve temel özellikleri

Türü	Temel Özellikler
Yatay – Dikey	Metrik ölçeklerin genel kullanımı yatay bir doğru parçası şeklindedir. Ancak literatürde, algı düzeyinin dikey bir doğru parçası vasıtasıyla daha hassas ölçülebileceği düşüncesiyle dikey metrik ölçekler geliştiren araştırmacılara rastlamak mümkündür (Dixon ve Bird, 1981).
Basit	Yatay ya da dikey bir doğru parçası üzerinde sadece uç değer etiketlerinin yer aldığı metrik ölçeğe basit metrik ölçek adı verilir.
Nümerik	Basit metrik ölçeğin, üzerinde sayısal etiketleri barındıran tipidir. Cevaplayıcının işaretlediği düzeyin sayısal olarak hangi dereceye karşılık geldiğini göstererek, cevaplayıcıyı yönlendirir.
Ortası İşaretli	Basit metrik ölçeğin ortasında işaret olan, böylece cevaplayıcının iki kutuplu doğru parçasının hangi tarafında daha yakın ise o tarafta işaretleme yapmasını kolaylaştıran bir tipidir (bu araştırmada ortası işaretli metrik ölçek kullanılmıştır).
Taksimli	Metrik ölçeğin belirli sayıda bölümlendirilmiş, örneğin beşli Likert tipi maddeler ile karşılaştırma yapılacak ise beş parçaya bölünmüş tipidir.
Sözel Etiketli	Taksimli metrik ölçeğin, üzerinde sözel derecelendirme etiketlerini bulduran (tamamen katılıyorum, katılmıyorum, kararsızım, vb.) tipidir.
Taksimli – Nümerik	Metrik ölçeğin belirli sayıda parçaya bölünmüş ve her parçası üzerinde sayısal etiketlerin bulunduğu tipidir.

Tablo 1’de verilen metrik ölçek türlerinden hangisinin kullanılacağına karar verirken araştırmanın amacı, madde sayısı, uygulama şekli (kâğıt – kalem / bilgisayar / tablet vb.), sanal platformlarda gerçekleştirilen uygulamalarda katılımcıların teknolojik cihaz kullanımında zorluk yaşamaması vb. hususlar dikkate alınmalıdır. Bu konularda karar vermede, metrik ölçeğin farklı yaş ve eğitim düzeylerine sahip bireylere uygulandığı çalışmalardan yararlanılabilir (Jensen, Karoly ve Braver, 1986; Laerhoven, Zaag-Loonen ve Derkx, 2004; Atan ve Yavuz, 2010).

Avantajlarının yanında Jensen, Karoly ve Braver (1986) metrik ölçeklerde iki hata kaynağından bahsetmektedir. Bu hatalardan ilki, cevaplayıcının söz konusu değişkene ilişkin durumu ile ölçek üzerindeki ilgili noktayı eşleştirmesi ve işaretlemesi sürecinden; ikincisi ise tepkinin veriye dönüştürülmesi aşamasındaki ölçme işleminden kaynaklanmaktadır.

Burada bahsedilen sorunlardan ilki kısmen de olsa Likert tipi ölçekler için de söz konusudur. Örneğin, bir cevaplayıcının “Katılmıyorum” şeklindeki tepkisinin, özelliğin hangi miktarına ya da düzeyine karşılık geldiği ve iki farklı maddeye verilen aynı “Katılıyorum” tepkisinin, özelliğin aynı miktarını temsil edip etmediği tam olarak bilinmemektedir. İkinci sorun için ise sıklıkla cetvel kullanılarak yapılan ölçme işlemi süresince dikkatli davranılması ve puanlama güvenilirliğini belirlemeye yönelik ek çalışmaların yapılması düşünülebilir. Cook vd. (2001) metrik ölçeklerin kâğıt – kalem uygulamalarında puanlamanın oldukça sıkıcı ve zaman alıcı olduğunu belirtmektedir. Ayrıca Snow ve Kirman (1988) fotokopi makinesi ayarlarının metrik ölçeğin uzunluğunu etkileyebileceğini belirtmekte, tüm kopyaların aynı formdan alınmasını ve uygulamadan önce ölçek uzunluğunun kontrol edilmesini önermektedir.

Uygulamaların bilgisayar ortamında gerçekleştirilmesi, cetvelle yapılan ölçme işlemi sürecindeki sorunları tümüyle ortadan kaldırılabilmektedir. Puanlamanın bilgisayarla yapılması ile zahmetsiz bir şekilde isabetli ölçümler elde edilebilmektedir. Fakat bu durumda da ölçeğin kaç piksele karşılık geleceği ve nasıl tasarlanacağı gibi yeni problemler ortaya çıkmaktadır. Reips ve Funke (2008) 50 – 800 piksel aralıklarında yürüttükleri çalışmada, aralarında önemli farklar olmasa bile 150 – 400 piksel aralığının kullanımını önermektedir.

Huskisson (1974) ağrının öznel bir şekilde ölçülmesinde en iyi sonuçların metrik ölçekler ile elde edilebileceğini belirtmektedir. Metrik ölçeklerin çeşitli türleri ağrı, yorgunluk, duyarlılık gibi değişkenlerin ölçülmesi amacıyla özellikle sağlık alanında yürütülen çalışmalarda sıklıkla kullanılmaktadır (Cengiz Çakır 2009; Cengiz, 2013; Akat, 2012; Carlsson, 1983). Metrik ölçeklerin eğitim ve psikoloji alanında yapılan çalışmalarda göreceli olarak daha az sayıda ele alındığı ve bu araştırmaların da genellikle güvenilirlik kapsamında yürütüldüğü görülmektedir.

Cook vd. (2001) kütüphane hizmet kalitesi ölçeğinin (LibQUAL+) bilgisayar ortamında dokuz dereceli ve metrik ölçek kullanılarak uygulanmasıyla elde edilen verilerini Cronbach α üzerinden incelemiştir. Metrik ölçek ile 0 – 100 aralığında elde edilen veriler daha sonra 1 – 5 ve 1 – 9 aralığına dönüştürülmüştür. Ölçeğin tüm alt boyutları için elde edilen değerler birbirine çok yakın olmakla beraber genel olarak en düşük güvenilirlik değerleri 1 – 5 aralığına dönüştürülen veri seti için elde edilmiş ve bunu sırasıyla, 1 – 9 aralığına dönüştürülen veri seti ve metrik ölçeğin izlediği tespit edilmiştir. En yüksek güvenilirlik değerleri dokuz dereceli formdan elde edilmiş olmakla beraber metrik ölçeğin internet üzerinden yürütülecek araştırmalarda güvenilir bilgiler verdiği belirtilmiştir.

Miller, Duncan, Brown, Sparks ve Claud (2003) üç faktörlü bir ölçek ile beraber bu ölçeğin faktörlerine ve geneline yönelik birer maddelik metrik ölçeğin yer aldığı formları kullanarak 45 maddelik dereceli bir ölçeği, dört maddelik metrik ölçek formuna dönüştürmeyi denemişlerdir. Her bir faktörün ve toplam puanın metrik ölçek ile korelasyonlarının .34 - .59; farklı uygulamalarla elde edilen test – tekrar test güvenilirliklerinin ise .49 - .66 arasında değiştiği, Cronbach α katsayısının ise .93 düzeyinde olduğu görülmüştür. Boer ve diğerleri (2004) de benzer şekilde genel yaşam kalitesinin tek maddelik metrik ölçek kullanılarak belirlenmesi ile elde edilen ölçümlerin test – tekrar test güvenilirliğinin .87; MOS SF-20 ve RSCL ölçekleri kullanılarak incelenen ölçüt bağımlı geçerliklerinin katsayılarının ise .29 - .72 arasında değiştiğini görmüşlerdir.

Gerich (2007) kâğıt – kalem ve bilgisayar ortamlarında uyguladığı öz kontrol ölçeğinin beş dereceli ölçek ile metrik ölçek verilerini karşılaştırmıştır. Metrik ölçek ve Likert ölçeği verileri üzerinden gerçekleştirilen temel bileşenler analizleri, öz kontrol ölçeğinin üç faktörlü yapısının her iki formda da ortaya konduğunu göstermiştir. İnternet üzerinden gerçekleştirilen uygulamalarda kâğıt – kalem uygulamasına göre daha yüksek ortalamalar elde edilmiş, kâğıt – kalem ortamındaki metrik ölçeğin güvenilirliğinin Likert ölçeğine göre, bilgisayar ortamında uygulanan Likert ölçeğinin güvenilirliğinin ise metrik ölçeğe göre daha yüksek olduğu belirlenmiştir. Ölçeklerin kâğıt – kalem uygulamasından elde edilen korelasyonlar, bilgisayar uygulamasından elde edilen korelasyonlara göre daha yüksektir.

Yıldız (1998) Likert tipi ve metrik ölçümlere dayalı olarak geliştirilen tutum ölçeğinin farklı formlarından elde edilen verilere dayalı olarak ölçeklerin psikometrik özelliklerini karşılaştırmıştır. Likert tipi ölçek ile elde edilen madde ayırt ediciliklerinin .54 – .88; metrik ölçekle elde edilen ayırt ediciliklerin .54 – .90 arasında değiştiği görülmüştür. %27’lik gruplar üzerinden yapılan karşılaştırmalarda metrik ölçeğin iki grubu daha iyi ayırt ettiği görülmüştür. Cronbach α katsayısı Likert tipi ölçek için .95 iken metrik ölçek için .96’dır. İki ölçekten elde edilen veriler arasındaki korelasyon katsayısı .94’tür.

Nartgün (2002) tarafından yürütülen çalışmada bir tutum ölçeğinin Likert tipi ve metrik formları ile elde edilen veriler KTK’ya ve MTK’ya göre incelenmiştir. KTK’ya göre yapılan incelemelerde her iki form için, madde ayırt ediciliklerinin birkaç

istisna haricinde .50'nin üzerinde olduğu ve ayırt edicilikler arasındaki korelasyonun .75 ($p < .01$); tutum ölçüleri arasındaki korelasyonun ise .91 ($p < .01$) düzeyinde olduğu görülmüştür. Matematikle İlgili Düşünceler Ölçeği'nin ölçüt geçerliği, fen ve sosyal alanlarda eğitim gören lise öğrencilerinin ortalamaları karşılaştırılarak incelenmiş ve her iki ölçekte de fen alanındaki öğrencilerin lehine manidar farklılıkların olduğu görülmüştür. Likert tipi ölçeğin Cronbach α katsayısı .96 iken metrik ölçeğin Cronbach α katsayısı .95'tir.

Koçyiğit (2002) tarafından yapılan çalışmada beş dereceli Likert tipi ölçek ile 10 ve 20 santimetre uzunluklarında metrik ölçekler karşılaştırılmıştır. Cronbach alfa katsayıları Likert tipi ölçek için .87, 10 cm'lik metrik ölçek için .91 ve 20 cm'lik ölçek için .90 düzeyindedir. Fisher z dönüşümü ile güvenilirlik değerleri arasında manidar fark bulunmamış olmakla birlikte en yüksek güvenilirlik değeri 10 cm'lik metrik ölçekle elde edilmiştir.

Kan (2009) Likert tipi, 0-100 ve metrik ölçek formatlarında hazırlana öz yeterlik ölçeğinin geçerlik ve güvenilirlik temelinde karşılaştırmıştır. Açıklayıcı faktör analizi sonuçlarına göre, üç veri seti için de tek faktörlü yapı elde edilmiş ve en yüksek açıklanan varyans oranı 0-100 ölçeği, en düşük açıklanan varyans oranı ise Likert tipi ölçekle elde edilmiştir. Cronbach α katsayısı ve madde ayırt edicilikleri incelendiğinde en yüksek değerlerin 0-100, en düşük değerlerin ise Likert tipi ölçek ile elde edildiği görülmüştür. KPSS ve akademik başarı puanlarını ölçüt olarak ele alındığı incelemelerde, 0-100 ölçeğinin diğer ölçek türlerinden daha iyi performans sergilediği görülmüştür.

Atan ve Yavuz (2010) aynı ölçeğin beş dereceli formu ile tablet üzerinden uygulanan metrik formundan elde edilen verileri çeşitli açılardan incelemiştir. Fisher z dönüşümü yapılarak karşılaştırılan Cronbach alfa katsayılarının metrik ölçek lehine daha yüksek olduğu ($z=1.702$, $p < .10$), ölçekteki on dört maddeden üçünün ayırt ediciliğinin metrik ölçek lehine daha yüksek olduğu görülmüştür. Likert tipi ve metrik ölçeklerden elde edilen toplam puanlar arasındaki korelasyon ise .74'tür. Katılımcıların %63,6'sı Likert tipi ölçeğin daha kolay cevaplandığını; %70,5'i ise metrik ölçek ile kendilerini daha iyi ifade ettiklerini belirtmiş ve bu durumun yaş ve eğitim düzeyine bağlı olmadığı belirlenmiştir.

Verilerin kâğıt – kalem formu ile ve internet üzerinden toplandığı karşılaştırmalı çalışmalarda, ölçme araçlarının psikometrik özellikleri arasında önemli farkların olmadığı ve bunula beraber uygulama süresi ve zamanı, puanlama, verilerin saklanması vb. konularda kullanışlı olduğu belirtilmektedir (Athale, Sturley, Skoczen, Kavanaugh ve Lenert, 2004; Coons ve diğerleri, 2009; Delgado ve diğerleri, 2018). Öte yandan Gül'ün (2012) çalışmasında internet üzerinden uygulanan testin güvenilirliğinin daha yüksek olduğu görülmüştür. Özel olarak, metrik ölçeklerin akıllı telefonlar kullanılarak cevaplanması ile yüksek güvenilirlik ve geçerliğin elde edildiği de bildirilmiştir (Caimmi ve diğerleri, 2017).

Yapılan araştırmalarda metrik ölçeklerden elde edilen verilerin psikometrik özelliklerinin sadece bir madde veya ölçek ile elde edildiği; ölçüt bağıntılı geçerliğe ilişkin incelemelerin de çeşitli dış ölçütlere göre ele alındığı görülmektedir. Bu araştırmada, Araştırmanın Yöntem Bölümüne İlişkin Öz Yeterlik Ölçeği (AYBİÖYÖ) ve Araştırmaya Yönelik Kaygı Ölçeği (AYKÖ) Likert tipi ve metrik ölçek formatında uygulanmış ve lisans örneklemindeki psikometrik özellikleri (faktör sayısı, açıklanan varyans oranı, iç tutarlılık katsayıları ve madde-toplam korelasyonları), kâğıt-kalem ve bilgisayar ortamındaki uygulamalara dayalı olarak incelenmiştir. Ölçeklerin ölçüt bağıntılı geçerliklerinin de yine bu veri setleri üzerinden incelenmiş olması, diğer bir deyişle ölçütün de metrik ölçek ile elde edilmiş olması bu araştırmanın önemini artırmaktadır.

2. YÖNTEM

Bu araştırma, AYBİÖYÖ'nün lisans düzeyindeki öğrencilere uygulanabilirliğinin incelendiği ve AYKÖ ile AYBİÖYÖ'nün psikometrik özelliklerinin farklı cevaplama ve uygulama koşullarında nasıl değiştiğini belirlemek amacıyla yürütülen ilişkisel türde bir araştırmadır.

2.1. Çalışma Grubu

Uygulamalar 2017 – 2018 Eğitim Öğretim Yılı Bahar Dönemi'nde Hasan Kalyoncu Üniversitesi ve Nevşehir Hacı Bektaş Veli Üniversitesi eğitim fakültelerindeki öğrenciler üzerinde gerçekleştirilmiştir. Öğrencilerin kayıtlı oldukları program ve sınıf düzeyine göre dağılımı Tablo 2'de verilmiştir.

Tablo 2.

Çalışma grubunda yer alan öğrencilerin program ve sınıf düzeyine göre dağılımı

Program	Sınıf				Toplam
	1	2	3	4	
Sınıf	33	32	5	42	112
PDR	1	52	2	57	112
Okul öncesi	24	-	1	-	25
İlköğretim matematik	44	1	2	-	47
Fen bilgisi	1	36	8	4	49
İngilizce	-	25	21	-	46
Toplam	103	146	39	103	391

Tablo 2’de program ve sınıf düzeyine göre dağılımı verilen 391 öğrenciden 261’i (%66,8) Hasana Kalyoncu Üniversitesi, 130’u (%33,2) Nevşehir Hacı Bektaş Veli Üniversitesi’nde öğrenim görmektedir. Öğrencilerden 316’sı kız (%80,4), 77’si (%19,6) erkektir, 200’ü (%51,2) bilimsel araştırma yöntemleri dersi almış 191’i (%48,8) almamıştır. AYBİÖYÖ’nün geçerliğine ek kanıt elde etmek amacıyla lisans ve lisansüstü düzeyde öğrenim görmekte olan öğrencilerin ortalamaları karşılaştırılmıştır. Çalışma grubunda, eğitim bilimleri alanında lisansüstü düzeyde öğrenime devam eden, Tablo 2’ye ve diğer analizlere dâhil edilmeyen 21 (4 erkek, 17 kız) öğrenci bulunmaktadır.

2.2. Veri Toplama Araçları

2.2.1. Araştırmaya Yönelik Kaygı Ölçeği (AYKÖ)

Üniversite öğrencilerinin araştırmaya yönelik kaygılarının belirlenmesinde kullanılmak amacıyla Büyüköztürk (1997) tarafından geliştirilmiştir. Çalışma grubunda 113 lisans ve 93 yüksek lisans öğrencisine yer verilen AYKÖ’nün kapsam geçerliği uzman görüşüne dayalı olarak incelenmiştir. Yapı geçerliği kapsamındaki incelemelerde tek faktör altında toplanan 12 maddenin toplam varyansın %41,9’unu açıkladığı ve faktör yüklerinin .55 - .74 arasında değiştiği görülmüştür. Ölçüt dayanaklı geçerlik kapsamında AYKÖ puanları ile Durumluk Kaygı Ölçeği (DKÖ) ve Sürekli Kaygı Ölçeği (SKÖ) puanları arasındaki ilişkiler incelenmiştir. AYKÖ ile DKÖ puanları arasında .30 ($p < .01$); SKÖ puanları arasında .51 ($p < .01$) düzeyinde ilişkilerin söz konusu olduğu görülmüştür. Yüksek lisans öğrencilerinin AYKÖ puanları ortalamasının lisans öğrencilerinin ortalamasına göre daha düşük olduğu görülmüştür. Beş (2, 3, 4, 8 ve 11) maddesi ters puanlanan AYKÖ’nün Cronbach α katsayısı .87’dir ve ölçekteki maddelerin ayırt edicilikleri .47 ile .65 arasında değişmektedir.

2.2.2. Araştırmanın Yöntem Bölümüne İlişkin Öz Yeterlik Ölçeği (AYBİÖYÖ)

Gök, Atalay Kabasakal ve Çetin (2015) tarafından, lisansüstü düzeyde öğrenim gören öğrencilerin araştırmaların yöntem bölümüne ilişkin öz yeterlik düzeylerinin belirlenmesinde kullanılmak üzere geliştirilmiştir. Lisansüstü eğitime devam eden 277 öğrenciden elde edilen veriler ölçeğin, 15 maddeden oluşan ve “veri analizi sırası ve sonrası (F1)” ve 13 maddeden oluşan “veri analizi öncesi (F2)” şeklinde adlandırılan iki faktörlü yapıya sahip olduğunu göstermiştir. Promax döndürme ile elde edilen ilk faktör toplam varyansın %47,60’ını, ikinci faktör ise %6,90’ını açıklamaktadır. Cronbach α katsayıları ilk faktör için .96, ikinci faktör için .88’dir. Madde - toplam korelasyonları .38 ile .86 arasında değişmektedir. İstatistik dersi alanların almayanlara, daha önce bir araştırma süreci deneyimi yaşamış olanların böyle bir süreçte bulunmamış olanlara ve doktora düzeyindeki öğrencilerin yüksek lisans ders ve tez dönemindeki öğrencilere göre manidar bir şekilde daha puanlar aldıkları belirlenmiştir.

2.3. Verilerin Toplanması

Eğitim fakültelerinde yürütülmekte olan ölçme ve değerlendirme, bilimsel araştırma yöntemleri, ölçek geliştirme gibi dersler kapsamında öğrenciler çeşitli araştırma süreçlerine yönlendirilebilmektedir. Özel olarak, rehberlik ve psikolojik danışmanlık (RPD) lisans programında yer alan seminer dersi kapsamında öğrencilerin ilgi ve ihtiyaçlarına göre RPD alanından seçtikleri konularda kapsamlı projeler hazırlamaları ve sunmaları beklenmektedir (YÖK, 2018). Öğrenciler, bu dersler kapsamında hazırladıkları araştırma raporlarını çeşitli kongrelerde sunabilmektedir (Ayaz, 2013; Demir, 2014; Osanmaz, Bilgi, İnce ve Tanju Aslışen, 2018). Bu durum göz önüne alınarak, AYBİÖYÖ’nün lisans düzeyinde de kullanılabileceği düşünülmüş ve bu araştırma kapsamında psikometrik özellikleri incelenmiştir.

Aun, Lam ve Collett (1986) tarafından Çinli hastalar üzerinde yürütülen bir araştırmada dikey metrik ölçeğin yatay forma göre daha az hata içerdiği ve bu durumun Çincenin yukarıdaki aşağıya ve sağdan sola okunan bir dil olmasından kaynaklanabileceği ifade edilmektedir. Hui ve Triandis (1989) tarafından gerçekleştirilen bir araştırmada da ölçek formunun kültüre ve etnik kökene göre farklılık gösterdiği belirtilmiştir. Ogon, Krismer, Söllner Kantner-Rumplmair ve Lampe (1996) yatay uygulanan ölçek ile elde edilen verilerin normal dağılım gösterdiğini fakat dikey formdan elde edilen verilerde normalliğin sağlanmadığını belirlemiştir. Burada belirtilen araştırmalara ve Türkçenin soldan sağa doğru yazılıp okunan bir dil olmasına dayanarak bu araştırmada metrik ölçekler yatay formatta uygulanmıştır.

Metrik ölçeğin kâğıt - kalem uygulamasında, 10 cm uzunluğunda ve iki ucu etiketlenmiş formların sıklıkla kullanıldığı ve önerildiği dikkate alınmıştır (Boer vd. 2004; Hawker, Mian, Kendzerska ve French, 2011; Harland, Dawkin ve Martin, 2015; Nartgün, 2002; Cook ve diğerleri, 2001; Koçyiğit 2002). Metrik ölçeğin internet üzerinden gerçekleştirilen uygulamalarında, iki ucun etiketlenmesi ve cevaplamaya başlandığında imlecin, ölçeğin orta noktasında yer alması önerilmektedir (Abend, Dan, Maoz, Raz ve Bar-Haim, 2014; Bosch, Revilla, DeCastellarnau ve Weber, 2018). Bu önerilere dayanarak bu araştırmada kâğıt - kalem ortamında kullanılan metrik ölçeğin 10 cm uzunluğunda olmasına, iki ucunun etiketlenmesine ve orta noktasının belirtilmesine karar verilmiştir. Benzer şekilde internet üzerinden gerçekleştirilen uygulamada da orta nokta belirtilmiştir. İnternette uygulanan metrik ölçeğin farklı tarayıcılarda aynı şekilde görünmesine dikkat edilmiş fakat cevaplayıcıların kullandığı cihazların özellikleri bilinemediği için uzunluğa ilişkin kesin bir belirleme yapılamamıştır.

İnternet üzerinden uygulanan metrik ölçek <http://kavramharitasiolustur.com/olcek/> adresinden yayınlanmıştır. Oluşturulan form sayfasının farklı cihazlar tarafından tutarlı bir şekilde çalışabilmesi için bootstrap kütüphanesi kullanılmıştır. Yayınlanan ekranda bulunan metrik ölçeğin örnek ekran görüntüsü Şekil 1’de verilmiştir.

#	AYBIÖYÖ	Hiç katılmıyorum	Tamamen katılıyorum
1	Araştırmanın yöntem bölümünü aşamalarına uygun olarak yazabilirim.	<input type="range" value="75"/>	
2	Araştırma yöntemini araştırma problemi ile tutarlı olacak şekilde belirleyebilirim.	<input type="range" value="75"/>	
3	Araştırma amacıma uygun araştırma desenini seçebilirim.	<input type="range" value="75"/>	
4	Araştırmada evren, örneklem ya da çalışma grubunu kolaylıkla belirleyebilirim.	<input type="range" value="75"/>	
5	Araştırma problemime uygun örnekleme yöntemini belirleyebilirim.	<input type="range" value="75"/>	
6	Araştırmamda kullanacağım örneklemin büyüklüğüne karar verebilirim.	<input type="range" value="75"/>	

Şekil 1. İnternet üzerinden uygulanan metrik ölçeğin örnek ekran görüntüsü

Şekil 1’de görüldüğü gibi ölçekte yer alan her bir madde karşısında kullanıcının belirli bir aralık belirtebileceği range adı verilen kaydırma çubukları bulunmaktadır. Range türündeki form kontrolleri HTML5 standartları sonrasında oluşturulmuş kontrollerdir. Çalışma kapsamında oluşturulan range kontrollerinin farklı cihazlarda (mobil cihazlar, bilgisayarlar, dokunmatik ekrana sahip cihazlar) uyumlu ve hatasız bir şekilde çalışabilmesi için rangeslider adında bir javascript kütüphanesinden yararlanılmıştır.

Kâğıt – kalem formları haftanın aynı günü, aynı saatte uygulanmıştır. Bu uygulamalarda öğrencilerden formların üzerine e-posta adreslerini yazmaları istenmiş ve böylece hem diğer ölçeklere ulaşmaları hem de verilerin eşleştirilmesi sağlanmıştır. Kâğıt – kalem uygulamalarından bir hafta sonra öğrencilere Likert tipi ölçeğe, bundan iki hafta sonra da aynı metrik ölçeğe erişim sağlayan bağlantılar gönderilmiştir. İnternet üzerinden yürütülen uygulamalar arasında iki haftalık zaman bırakılmasının sebebi, birer hafta sonunda katılım göstermeyenlere hatırlatma e-postasının gönderilmesidir.

Son hatırlamadan bir hafta sonra internet üzerinden uygulanan formlardan elde edilen veriler bilgisayar ortamına aktarılmıştır. Kâğıt – kalem formu için metrik ölçeği cevaplayan öğrenci sayısı 298, Likert tipi ölçeği cevaplayan öğrenci sayısı 282’dir. İnternet üzerinden yürütülen uygulamada ise metrik ölçeği cevaplayan öğrenci sayısı 152, Likert tipi ölçeği cevaplayan öğrenci sayısı ise 85’tir. Uygulamaların dördüne de katılan öğrenci sayısı ise 33’tür. Likert tipi ölçeği cevaplayan katılımcılardan 77’si cep telefonu, sekizi bilgisayar; metrik ölçeği cevaplayan katılımcılardan ise 111’i cep telefonu 41’i bilgisayar kullandığını belirtmiştir.

2.4. Verilerin Analizi

Ölçeklerin psikometrik özellikleri temel bileşenler analizi, Cronbach α ve eşdeğer yarılar güvenilirliği ve madde – toplam korelasyonları üzerinden incelenmiştir. Ölçeklerden alınan puanlar arasındaki ilişkilerin belirlenmesinde Pearson korelasyon katsayısı kullanılmış, güvenilirlikler arasındaki farkların manidarlığı Fisher z ile test edilmiştir. Lisansüstü düzeyde eğitim gören bireylere uygulanmak üzere geliştirilmiş olan AYBIÖYÖ’nün lisans düzeyindeki öğrencilere uygulanmasıyla elde edilen puanlarının geçerliliği hakkında bilgi edinmek amacıyla, puanların bilimsel araştırma yöntemleri dersini alıp almama ve öğrenim düzeyine göre manidar fark gösterip göstermediği ilişkisiz ölçümler için t testi ile incelenmiştir.

Kâğıt – kalem formundaki metrik ölçekler iki araştırmacı tarafından aynı marka cetvel kullanılarak puanlanmıştır. Rastgele seçilen 50 formdaki cevaplar farklı bir puanlayıcı tarafından tekrar ölçülmüş ve puanlama güvenilirliğine yönelik incelemeler Pearson korelasyon katsayısı üzerinden gerçekleştirilmiştir.

Analizlerden önce farklı veri setlerinde %0,86 - %1,63 oranlarında olduğu tespit edilen kayıp verilerin seçkisizliği Little’s MCAR testine ek olarak kayıpların kişi ve maddelere dağılımı üzerinden incelenmiş ve tümüyle seçkisiz kayıp örüntüsünde olduğu tespit edilerek beklenti – maksimizasyon algoritması ile veri setleri tamamlanmıştır (Akbaş ve Tavşancıl, 2015). Standart z puanı [-3, +3] aralığı dışında kaldığı belirlenen ve farklı veri setlerinde sayıları bir ile altı arasında değişen uç değerlerin ilgili büyüklükteki örneklem için kabul edilebilir bir durum olduğu göz önüne alınarak veri setinde kalmalarına karar verilmiştir.

AYKÖ'de yer alan beş madde (2, 3, 4, 8 ve 11) ters kodlanmaktadır. Kâğıt – kalem formunda metrik ölçeğin kullanıldığı uygulamalarda bu maddelere verilen tepkiler 10'dan çıkarılmıştır. İnternet üzerinden yürütülen uygulamalarda ise puanlama, ölçeğin sol ucu 10 sağ ucu sıfır olacak şekilde belirlenmiştir.

3. BULGULAR

Öncelikle farklı formların uygulanması ile elde edilen veriler üzerinden hesaplanan betimsel istatistiklere, ölçeklerin psikometrik özelliklerine ilişkin inceleme sonuçlarına yer verilmiştir. Farklı uygulamalardan elde edilen betimsel istatistikler Tablo 3'te sunulmuştur.

Tablo 3.

Ölçeklerin farklı uygulamalarından elde edilen betimsel istatistikler

Ölçme aracı ve uygulama şekli			N	Ortalama	St. sapma	En k.	En b.	Bas.	Çarp.
AYBİÖYÖ	Kağıt-kalem	Likert	282	93.98	16.58	33	136	1.21	-.81
		Metrik	298	131.49	42.59	6.65	231.10	.30	-.48
	İnternet	Likert	85	89.75	24.94	28	140	.59	-.58
		Metrik	152	152.35	46.43	12	253.80	.20	-.58
AYKÖ	Kağıt-kalem	Likert	282	29.53	8.72	12	59	.22	.49
		Metrik	298	44.62	23.20	0	119.80	.45	.63
	İnternet	Likert	85	27.16	8.62	12	49	-.64	.16
		Metrik	152	40.26	23.18	0	108.80	-.36	.28

Tablo 3'te aynı cevaplama formatından elde edilen verilere ait eğilim ve dağılım ölçülerini birbirine yakın olduğu, basıklık ve çarpıklık değerlerine göre de toplam puanların dağılımının normalden önemli bir sapma göstermediği görülmektedir.

3.1. AYBİÖYÖ'nün psikometrik özelliklerine ilişkin bulgular

AYBİÖYÖ'nün yapı geçerliğinin, farklı ölçek türleri ve farklı uygulama koşullarında nasıl değiştiğini belirlemek amacıyla gerçekleştirilen temel bileşenler analizlerine ilişkin sonuçlar Tablo 4'te verilmiştir.

Tablo 4.

Farklı form ve ortamlarda elde edilen veriler için temel bileşenler analizi sonuçları

Maddeler	Kâğıt - Kalem						İnternet			
	Özgün FY		Likert FY		Metrik FY		Likert FY		Metrik FY	
	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
M1	.44		.59	.06	.55	.25	.65	.20	.50	.27
M2	.35		.47	.19	.54	.25	.49	.43	.43	.37
M3	.41		.70	-.01	.46	.38	.66	.23	.50	.25
M4	.42		.73	.02	.50	.35	.66	.20	.61	.19
M5	.78		.76	-.01	.74	.11	.66	.24	.84	-.03
M6	.89		.81	-.14	.91	-.15	.94	-.10	.85	-.15
M7	.88		.80	-.07	.76	.11	.84	-.01	.82	-.01
M8	.95		.76	-.02	.91	-.06	1.02	-.20	.82	-.04
M9	.92		.75	-.06	.87	-.07	1.06	-.25	.86	-.16
M10	.95		.77	-.01	.80	-.01	.86	.05	.81	.02
M11	.93		.71	-.02	.78	.07	.82	.06	.80	-.02
M12	.36		.55	.26	.87	-.02	.83	.05	.69	.11
M13	.85		.74	.03	.82	.02	.80	.08	.56	.25
M14	.87		.78	-.10	.83	-.11	.74	.09	.44	.23
M15	.59		.65	.07	.63	.13	.56	.31	.82	-.11
M16		.46	-.04	.76	-.01	.75	-.17	1.00	-.22	.93
M17		.49	-.15	.77	-.05	.87	-.01	.90	.04	.77
M18		.49	-.02	.71	-.06	.86	.08	.68	.03	.70
M19		.73	-.04	.75	-.14	.92	-.15	.98	-.20	.96
M20		.58	-.09	.79	-.08	.87	.08	.80	.02	.81
M21		.50	-.08	.79	-.06	.84	-.20	.99	-.08	.86
M22		.60	.09	.64	.11	.67	.10	.78	.17	.61
M23		.61	.20	.56	.24	.63	.23	.67	.25	.57
M24		.65	.29	.40	.08	.74	.19	.70	.08	.70
M25		.55	.43	.10	.44	.20	.40	.33	.37	.24
M26		.56	.34	.44	.22	.55	.16	.70	.28	.52
M27		.62	.33	.44	.28	.56	.19	.70	.06	.71
M28		.63	.32	.28	.15	.52	.36	.43	.32	.41
N		277		282		298		85		152
KMO		.94		.94		.95		.92		.92
Bartlett χ^2		7009.29**		4350.95**		7332.71**		2698.94**		3296.36**
Öz değerler		13.36 (1.75) ^a		11.732 (1.64) ^a		15.36 (1.62) ^a 1.97 (1.52) ^a		17.44 (2.18) ^a 2.28 (2.00) ^a		13.82 (1.90) ^a 2.19 (1.75) ^a
		2.06 (1.63) ^a		2.07 (1.54) ^a		1.23 (1.46) ^a		1.43 (1.84) ^a		1.38 (1.65) ^a
		1.59 (1.55) ^a		1.44 (1.47) ^a						
A. Varyans		%47,60		%41,90		%54,85		%62,30		%49,34
		%6,90		%7,40		%7,04		%8,16		%7,82
Toplam A. V.		%54,50		%49,30		%61,90		%70,46		%57,16

**p<.01, a Paralel analiz ile elde edilen öz değerler

Tablo 4'te dört farklı uygulamadan elde edilen veriler üzerinden gerçekleştirilen temel bileşenler analizi sonuçları görülmektedir. KMO değerlerinin .92 - .95 aralığında değiştiği ve tüm Bartlett testlerinin manidar olduğu görülmektedir (p<.01). Tüm veri setlerinde özgün ölçeğin iki faktörlü yapısı, elde edilen öz değerler ve paralel analiz sonuçları ile desteklenmektedir. Açıklanan varyans oranları incelendiğinde, en düşük açıklanan toplam varyans oranının %49,30 ile kâğıt - kalem formatında uygulanan Likert tipi ölçek ile elde edildiği görülmektedir. Bunu sırasıyla internet üzerinden uygulanan metrik ölçek (%57,16) ve kâğıt - kalem formatındaki metrik ölçek (%61,90) izlemektedir. En yüksek açıklanan varyans oranı ise internet ortamında uygulanan Likert tipi ölçek ile elde edilmiştir.

Özgün ölçekte olduğu gibi promax eğik döndürme sonucunda elde edilen faktör yükleri incelendiğinde, üç farklı veri setinde M25'in tutarlı bir şekilde birinci faktör altında .40'ın üzerinde yüke sahip olduğu görülmektedir. İnternet üzerinden metrik ölçeğin uygulanmasıyla elde edilen değerler incelendiğinde de bu maddenin .40'ın bir miktar altında olmakla birlikte, birinci faktördeki yükünün daha yüksek olduğu görülmektedir. M28'in kâğıt - kalem formatındaki Likert tipi ölçek ile elde edilen faktör yüklerinin .40'ın altında olduğu fakat benzer bir durumun diğer veri setlerinde söz konusu olmadığı görülmektedir. M8, M9 ve M16'nın internet üzerinden uygulanan Likert tipi ölçekten elde edilen veriler için faktör yüklerinin birin üzerinde olması ise eksenlerin eğik döndürüldüğü durumlar için olası bir durumdur (Jöreskog, 1999).

Birinci faktör altında yüksek yüke sahip olan M25 (Başka kültürlerde geliştirilen ölçme araçlarını kendi kültürümüze / dilimize uyarlayabilirim) incelendiğinde, birinci faktörde yer alan M2'ye (Ölçme aracı geliştirmek için yapılması gereken tüm

işlemleri / aşamaları yerine getirebilirim) kavramsal olarak yakın olduğu söylenebilir. M2 ve M25'te ifade edilen süreçlerin her ikisinin sonucunda bir ölçme aracına ulaşılmaktadır. Bu durum AYBİÖYÖ'yü geliştiren araştırmacılar ile paylaşılmış ve ölçeğin lisans düzeyinde uygulanması halinde M25'in birinci faktör altında değerlendirilmesinin kabul edilebilir bir durum olduğu sonucuna ulaşılmıştır. Yapılan değişikliğe bağlı olarak da bu noktadan sonra, M25'e birinci faktör altında yer verilerek elde edilen bulgular (faktör puanları, iç tutarlılık katsayıları vb.) sunulmuştur. M28'in .40'ın altında ve M2'nin her iki faktörde de .40'ın üzerinde faktör yüküne sahip olduğu durumlar ise istisna olarak değerlendirilmiş ve herhangi bir değişikliğe gidilmemiştir.

AYBİÖYÖ'de yer alan maddelerin ayırt edicilikleri düzeltilmiş madde – toplam korelasyonu ile güvenilirliği ise Cronbach α ve eşdeğer yarılar yöntemleri ile incelenmiştir. Elde edilen sonuçlar Tablo 5'te verilmiştir.

Tablo 5.
AYBİÖYÖ'nün güvenilirliği ve maddelerin ayırt edicilikleri

Uygulama şekli ve ölçek türü		Faktör	rjx	Cronbach α	Eşdeğer yarı
Kâğıt – kalem	Likert	F1	.47 - .66	.90	.78
	Likert	F2	.46 - .74	.93	.88
	Likert	Toplam	.47 - .72	.95	.87
	Metrik	F1	.59 - .80	.94	.87
	Metrik	F2	.56 - .82	.96	.92
	Metrik	Toplam	.57 - .78	.97	.91
İnternet	Likert	F1	.67 - .86	.96	.88
	Likert	F2	.63 - .87	.97	.92
	Likert	Toplam	.64 - .84	.98	.90
	Metrik	F1	.61 - .77	.93	.86
	Metrik	F2	.53 - .77	.94	.92
	Metrik	Toplam	.54 - .73	.96	.87

Tablo 5'te, tüm uygulamalar için elde edilen madde toplam korelasyonlarının .46 - .87 aralığında değiştiği ve madde ayırt ediciliklerinin yüksek olduğu görülmektedir. Güvenirlik katsayıları incelendiğinde de göreceli olarak en düşük değerlerin söz konusu olduğu kâğıt – kalem formatındaki Likert tipi ölçek için elde edilen katsayıların oldukça yüksek olduğu anlaşılmaktadır. Bununla beraber en yüksek ayırt edicilik ve güvenirlik değerlerinin internet üzerinden uygulanan Likert tipi ölçekle elde edildiği söylenebilir. Ölçek genelinden elde edilen güvenirlik katsayıları arasındaki farkın manidarlığa ilişkin Fisher z testi sonuçları Tablo 6'da verilmiştir.

Tablo 6.
AYBİÖYÖ'nün güvenilirliğine ilişkin Fisher z testi sonuçları

Uygulama şekli ve ölçek türü		Kâğıt-kalem		İnternet	
		Likert	Metrik	Likert	Metrik
Kâğıt-kalem	Likert	-	2.33*	1.11	.00
	Metrik	3.11**	-	.45	1.94
İnternet	Likert	3.71**	1.65	-	1.01
	Metrik	1.12	1.45	2.56*	-

Tablo 6'nın üst köşegeninde koyu yazılan değerler eşdeğer yarılar, alt köşegenindeki değerler ise Cronbach α katsayıları arasındaki farkın manidarlığı için hesaplanan z değerleridir. Eşdeğer yarılar güvenilirliği için, kâğıt – kalem ortamında uygulanan metrik ölçeğin güvenilirliği Likert tipi ölçeğe göre daha yüksektir ($p < .05$). Kâğıt – kalem ortamındaki Likert tipi ölçeğin güvenilirliği metrik ölçeğe ($p < .01$) ve internet uygulamasındaki Likert ölçeğine ($p < .01$) göre düşük iken internet üzerinden uygulanan metrik ölçek ile aralarında manidar fark yoktur ($p > .05$). Ayrıca internet üzerinden uygulanan İnternet üzerinden yürütülen uygulamalarda Likert ölçeğinin metrik ölçeğe göre daha güvenilir olduğu anlaşılmaktadır ($p < .05$).

Bilimsel araştırma yöntemleri (BAY) dersi almış ve almamış olan lisans öğrencilerinin AYBİÖYÖ faktörlerinden ve toplamından aldıkları puanlar arasındaki farkların manidarlığı ilişkisiz ölçümler için t testi ile incelenmiştir. Elde edilen sonuçlar Tablo 7'de yer almaktadır (sadece kâğıt – kalem formunda uygulanan Likert tipi ölçekten elde edilen sonuçlar verilmiştir).

Tablo 7.
AYBİÖYÖ puanlarının bilimsel araştırma yöntemleri dersi alma göre karşılaştırılmasına ilişkin t testi sonuçları

Faktör	BAY Dersi	N	Ortalama	Standart sapma	sd	t	p	Cohen d
F1	Alan	141	54.24	8.68	264.01	3.30	.00	.39
	Almayan	141	50.31	11.16				
F2	Alan	141	45.09	5.54	248.09	8.21	.00	.98
	Almayan	141	38.32	8.07				
Toplam	Alan	141	99.33	13.28	258.79	5.72	.00	.68
	Almayan	141	88.63	17.82				

Tablo 7’de, bilimsel araştırma yöntemleri dersini almış olan öğrencilerin AYBİÖYÖ faktör ve toplam puanları ortalamalarının bu dersi almamış öğrencilere göre manidar bir şekilde daha yüksek olduğu görülmektedir ($p<.01$). Etki büyüklüğüne ilişkin Cohen d değerleri veri analizi sırası ve sonrası faktörü (F1) için küçük, veri analizi öncesi faktörü (F2) için geniş ve toplam puanlar için ise orta düzey bir etkinin söz konusu olduğu anlaşılmaktadır (Büyüköztürk, 2018). AYBİÖYÖ faktör ve toplam puanlarının eğitim düzeyine göre karşılaştırılmasına ilişkin sonuçlar Tablo 8’de verilmiştir.

Tablo 8.

AYBİÖYÖ puanlarının öğrenim düzeyine göre karşılaştırılmasına ilişkin t testi sonuçları

Faktör	Öğrenim	N	Ortalama	Standart sapma	sd	t	p	Cohen d
F1	Lisans	282	52.27	10.18	301	2.43	.02	.55
	Lisansüstü	21	57.86	9.93				
F2	Lisans	282	41.71	7.70	301	4.77	.00	1.08
	Lisansüstü	21	49.90	6.09				
Toplam	Lisans	282	93.98	16.58	301	3.70	.00	.84
	Lisansüstü	21	107.76	14.98				

Tablo 8’de, AYBİÖYÖ faktör ve toplam puan ortalamalarının lisansüstü düzeyde öğrenim görmekte olan öğrenciler lehine manidar bir şekilde daha yüksek olduğu görülmektedir ($p<.05$). Cohen d değerleri incelendiğinde de “veri analizi sırası ve sonrası” faktörü için orta (.55), “veri analizi öncesi” faktörü (1.08) ve ölçek geneli (.84) için geniş düzeyde bir etki büyüklüğünün söz konusu olduğu görülmektedir.

3.2. AYKÖ’nün psikometrik özelliklerine ilişkin bulgular

AYKÖ’nün yapı geçerliğinin, farklı ölçek türleri ve farklı uygulama koşullarında nasıl değiştiğini belirlemek amacıyla gerçekleştirilen temel bileşenler analizlerine ilişkin sonuçlar Tablo 10’da verilmiştir.

Tablo 9.

Farklı form ve ortamlarda elde edilen veriler için temel bileşenler analizi sonuçları

Maddeler	Kâğıt - Kalem			İnternet	
	Özgün FY	Likert FY	Metrik FY	Likert FY	Metrik FY
M1	.64	.72	.74	.71	.75
M2	.55	.55	.74	.78	.76
M3	.64	.70	.77	.83	.80
M4	.58	.68	.75	.66	.56
M5	.71	.79	.63	.82	.87
M6	.68	.77	.82	.79	.76
M7	.70	.80	.82	.77	.79
M8	.72	.68	.79	.71	.76
M9	.74	.79	.84	.79	.82
M10	.63	.66	.71	.66	.81
M11	.57	.26	.41	.41	.38
M12	.59	.67	.70	.62	.66
N	202	282	298	85	152
KMO	-	.91	.93	.90	.91
Bartlett χ^2	-	1579.81**	2491.13**	677.51**	1228.95**
Öz değerler	-	5.68 (1.35) ^a	6.74 (1.34) ^a	6.26 (1.67) ^a	6.55 (1.48) ^a
	-	1.27 (1.26) ^a	1.36 (1.25) ^a	1.73 (1.48) ^a	1.58 (1.35) ^a
Aç. Varyans	%41,90	%47,34	%56,13	%52,17	%54,55

** $p<.01$, a Paralel analiz ile elde edilen öz değerler

Tablo 9’da bütün veri setleri için elde edilen KMO değerlerinin .90 - .93 aralığında ve Bartlett testlerini manidar olduğu ($p<.01$) ve farklı veri setleri için 1’den büyük öz değer sayısının iki olduğu görülmektedir. Bu öz değerler ilgili örneklem büyüklükleri için paralel analiz (Watkins, 2000) sonuçları ile karşılaştırıldığında iki faktörlü yapının desteklenebileceği anlaşılmaktadır. İlk öz değer ikinci öz değere göre büyük olması, ikinci öz değer 1’e ve paralel analiz ile elde edilen değere yakın olması, ilk faktörün açıkladığı varyans oranının bir faktör için yüksek sayılabileceği (Büyüköztürk, 2018) ve AYKÖ’nün özgün yapısının bir faktörlü olması gerekçeleriyle analizler bir faktöre göre gerçekleştirilmiştir.

Faktör yüklerinin, kâğıt - kalem formundaki Likert tipi ölçek için .26 - .80, metrik ölçek için .41 - .84; internet üzerinden gerçekleştirilen uygulamalar için ise sırasıyla .41 - .83 ve .38 - .87 aralığında değiştiği ve tüm koşullarda m11’in en düşük faktör yüküne sahip olduğu görülmektedir. En düşük açıklanan varyans oranı Likert tipi ölçeğin kâğıt - kalem formundan elde edilirken bunu sırasıyla internet üzerinden uygulanan Likert tipi ölçek ve metrik ölçek izlemektedir. En yüksek açıklanan varyans oranına kâğıt - kalem ortamında uygulanan metrik ölçek ile ulaşılmıştır.

AYKÖ'de yer alan maddelerin ayırt edicilikleri düzeltilmiş madde – toplam korelasyonu ile güvenilirliği ise Cronbach α ve eşdeğer yarılar yöntemleri ile incelenmiştir. Elde edilen sonuçlar Tablo 10'da verilmiştir.

Tablo 10.

Farklı form ve ortamlarda elde edilen madde ayırt edicilik ve güvenilirlik katsayıları

Maddeler	Kâğıt - Kalem			İnternet	
	Özgün rjx	Likert rjx	Metrik rjx	Likert rjx	Metrik rjx
M1	.55	.65	.68	.66	.67
M2	.47	.49	.69	.74	.73
M3	.55	.63	.72	.80	.76
M4	.50	.62	.70	.60	.52
M5	.62	.72	.74	.76	.81
M6	.58	.69	.77	.71	.70
M7	.60	.72	.76	.69	.72
M8	.64	.62	.75	.66	.72
M9	.65	.71	.79	.71	.75
M10	.54	.58	.65	.59	.74
M11	.49	.23	.36	.36	.34
M12	.50	.58	.63	.54	.59
N	202	282	298	85	152
Cronbach α	.87	.89	.93	.91	.92
Eşdeğer yarılar	-	.88	.92	.93	.93

Tablo 10'da belirtilen ayırt edicilik değerlerinin kâğıt – kalem formundaki Likert tipi ölçek için .23 - .72, metrik ölçek için .36 - .79; internet ortamında gerçekleştirilen uygulamalar için ise sırasıyla .36 - .80 ve .34 - .81 aralığında değiştiği ve tüm koşullarda m11'in en düşük ayırt ediciliğe sahip olduğu görülmektedir. Farklı veri setleri için elde edilen Cronbach α katsayılarının .89 - .93, eşdeğer yarılar güvenilirliklerinin ise .88 - .93 aralığında değiştiği görülmektedir. Cronbach α ve eşdeğer yarılar güvenilirlik katsayıları arasındaki farkların manidarlığı Fisher z testi ile incelenmiştir. Sonuçlar Tablo 11'de belirtilmiştir.

Tablo 11.

AYKÖ'nün güvenilirliğine ilişkin Fisher z testi sonuçları

Uygulama şekli ve ölçek türü		Kâğıt-kalem		İnternet	
		Likert	Metrik	Likert	Metrik
Kâğıt-kalem	Likert	-	2.55*	2.24*	2.78**
	Metrik	2.82**	-	.55	.69
İnternet	Likert	.84	1.04	-	.00
	Metrik	1.64*	.69	.44	-

*p<.05, **p<.01

Tablo 11'in üst köşegeninde koyu yazılan değerler eşdeğer yarılar, alt köşegenindeki değerler ise Cronbach α katsayıları arasındaki farkın manidarlığı için hesaplanan z değerleridir. Eşdeğer yarılar güvenilirliği için, kâğıt – kalem ortamında uygulanan Likert türü ölçekle elde edilen güvenilirlik değerlerinin diğer uygulamalardan elde edilenlere göre daha düşük olduğu anlaşılmaktadır (p<.05). Cronbach α katsayıları için elde edilen değerlere göre ise metrik ölçeklerle elde edilen ölçümlerin güvenilirliği kâğıt – kalem formundaki Likert ölçeğine göre daha yüksektir (p<.05). Diğer güvenilirlik katsayıları arasında manidar bir fark söz konusu değildir (p>.05).

AYBİÖYÖ ve AYKÖ'nün ölçüt bağıntılı geçerliği, farklı ortamlarda ve formlarda gerçekleştirilen uygulama verileri üzerinden korelasyon analizi ile incelenmiştir. AYBİÖYÖ ve AYKÖ toplam puanları arasındaki korelasyon katsayıları Tablo 12'de verilmiştir.

Tablo 12.
AYBİÖYÖ ve AYKÖ puanları arasındaki korelasyonlar

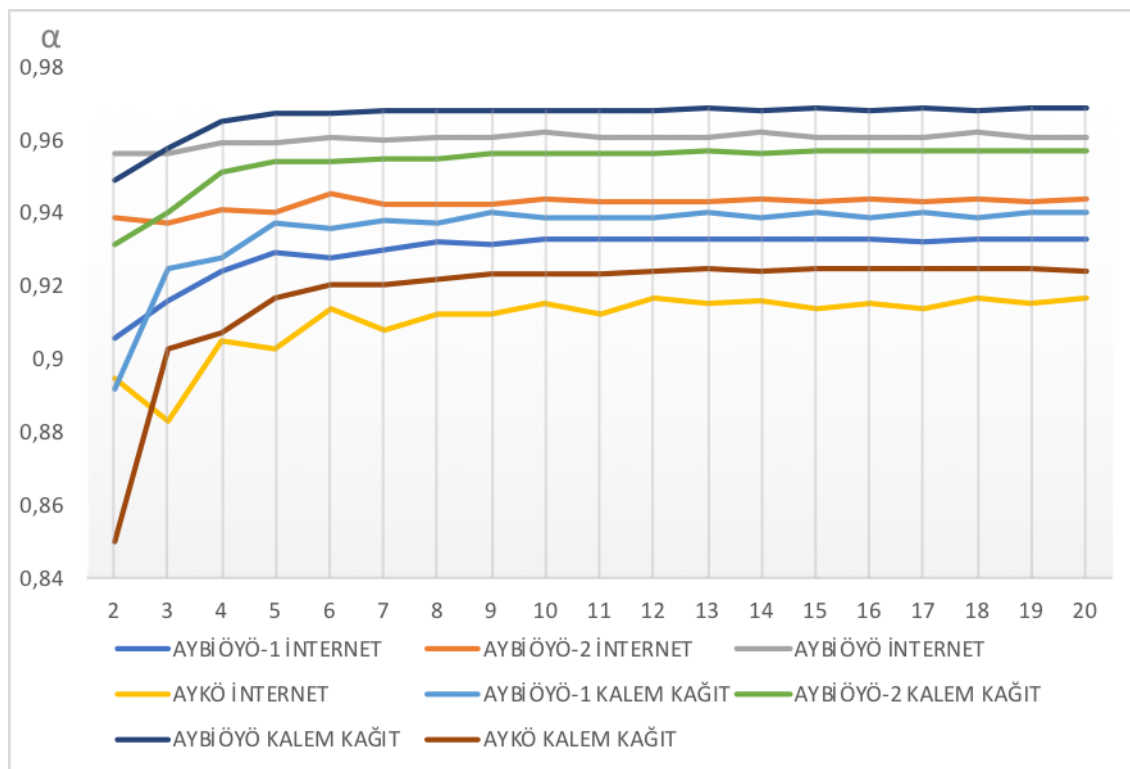
Ölçek	Uygulama	1	2	3	4	5	6	7	8
AYBİÖYÖ	Kâğıt kalem	1. Likert	-						
		2. Metrik	.85**	-					
	İnternet	3. Likert	.81**	.75**	-				
		4. Metrik	.38*	.58**	.46**	-			
AYKÖ	Kâğıt kalem	5. Likert	-.58**	-.60**	-.36*	-.38*	-		
		6. Metrik	-.40*	-.50**	-.18	-.35*	.81**	-	
	İnternet	7. Likert	-.44*	-.50**	-.36*	-.30	.78**	.76**	-
		8. Metrik	-.52**	-.52**	-.32	-.20	.78**	.79**	.77**

*p<.05, **p<.01

Tablo 12’de tüm uygulamalara eksiksiz bir şekilde katılan 33 öğrenciden elde edilen puanlar arasındaki korelasyonlar yer almaktadır. Elde edilen korelasyonlar genel olarak değerlendirildiğinde, boyalı bir şekilde belirtilen hücrelerdeki değerler ölçüt geçerliği bağlamında diğer korelasyonlara göre göreceli olarak daha yüksek sayılabilir. Buna göre, AYBİÖYÖ’nün tüm uygulamaları için, AYKÖ ile en yüksek korelasyonların elde edildiği durum, AYKÖ’nün kâğıt – kalem formatında Likert tipi ölçek kullanılarak uygulandığı durumdur. AYKÖ için ise en yüksek korelasyonlar, AYBİÖYÖ’nün kâğıt – kalem formatında metrik ölçek kullanılarak uygulanmasıyla elde edilmiştir.

Puanlama güvenilirliği kapsamındaki incelemelerde, farklı bir puanlayıcının metrik ölçeğin uygulandığı formlar arasından seçkisiz olarak belirlediği 50 formu, aynı cetveli kullanarak tekrar puanlanması istenmiştir. Bağımsız puanlayıcılardan elde edilen veriler üzerinden madde çiftleri ve ölçek toplam puanları arasındaki korelasyonlar incelenmiştir. İncelemeler sonucunda AYBİÖYÖ’de bulunan maddeler için .90 – 1.00; AYKÖ’de bulunan maddeler için ise .98 – 1.00 arasında korelasyonların bulunduğu görülmüştür. Her iki ölçek için toplam puanlar arasındaki korelasyon 1’dir.

Her ne kadar araştırmamızın temel amaçları arasında yer almasa da Likert tipi ölçeklerde yer verilmesi önerilen kategori sayısı konusundaki tartışmalara katkıda bulunmak amacıyla, farklı ortamlarda uygulanan metrik ile elde edilen ölçümler, aralıkları eşit olarak şekilde 2 – 10 kategoriye ayrılmış ve her bir veri seti için Cronbach α değerleri hesaplanmıştır. Farklı kategori sayıları için elde edilen değerler Şekil 2’de verilmiştir.



Şekil 2. Metrik ölçeklerin farklı sayıda kategorilere ayrılmasıyla elde edilen Cronbach α katsayıları

Şekil 2’de, Cronbach α katsayılarının kategori sayısı ile birlikte artma eğiliminde olduğu ve en büyük artışın 2-3-4 kategori arasında yaşandığı görülmektedir. Artışın genel olarak beş kategoriden itibaren yavaşladığı ve yedi kategoriden sonra önemsiz sayılabileceği söylenebilir.

4. TARTIŞMA ve SONUÇ

İki ölçeğin cevaplama formatı ve cevaplama ortamının manipüle edildiği bu çalışmada öncelikle, lisansüstü düzeyde öğrenim gören öğrenciler için geliştirilmiş olan AYBİÖYÖ’nün psikometrik özellikleri, lisans düzeyindeki öğrencilerden elde edilen veriler üzerinden incelenmiştir. Dört farklı veri seti üzerinden gerçekleştirilen temel bileşenler analizi sonuçları, özgün formda ikinci faktörde yer alan bir maddenin (M25) birinci faktör altında yer alması dışında aynı yapının ortaya konduğunu göstermiştir.

Güvenirlilik temelinde yürütülen incelemelerde Cronbach α ve eşdeğer yarı güvenirliliklerinin ölçeği oluşturan faktörler ve tüm ölçek için .78 - .96 aralığında olduğu ve madde - toplam korelasyonlarının tüm veri setleri için .47 - .86 arasında değiştiği görülmüştür. Bilimsel araştırma dersi almış olan öğrencilerin faktör ve toplam puan ortalamaları bu dersi almayanlara göre; öğrenimlerine lisansüstü düzeyde devam eden öğrencilerin de lisans düzeyindekilere göre manidar bir şekilde yüksektir. Öz yeterlik ile kaygı ve araştırma yeterliği ve araştırma kaygısı yapıları arasında (-.16 ile -.65 arasında değişen) negatif ve zayıf/orta düzey korelasyonların bulunduğu bildirilmektedir (Saracaloğlu, 2008; Yıldırım, 2011; Maharajan ve diğerleri, 2017). Benzer şekilde bu çalışmada da lisans düzeyindeki öğrencilerin araştırma kaygısı ve araştırmaların yöntem bölümüne ilişkin öz yeterlikleri arasında negatif ve zayıf / orta düzey korelasyonlara ulaşılmıştır.

Elde edilen bu bulgular bir bütün olarak değerlendirildiğinde Gök, Atalay Kabasakal ve Çetin (2015) tarafından geliştirilen AYBİÖYÖ’nün, lisans düzeyinde öğrenim görmekte olan öğrencilerin araştırmanın yöntem bölümüne ilişkin öz yeterliklerinin ölçülmesi amacıyla kullanılmaya uygun psikometrik özelliklere sahip olduğu anlaşılmaktadır. Lisans düzeyindeki öğrencilere uygulandığında, özgün formda “veri analizi öncesi” faktöründe yer alan M25’in “veri analiz sırası ve sonrası” faktöründe yer almasına dikkat edilmelidir.

Farklı ortamlarda uygulanan ölçeklere ait betimsel istatistikler incelendiğinde Gerich (2007)’ye benzer şekilde internet uygulamalarında daha yüksek aritmetik ortalamaların elde edildiği görülmüştür. Bu çalışmada dikey metrik ölçek karşılaştırması yapılmamış olmakla birlikte yatay metrik ölçekle elde edilen çarpıklık ve basıklık katsayıları Hui ve Triandis (1989) ve Ogon ve diğerlerinin (1996) ifadelerine benzer şekilde normal dağılımın sağlandığını ortaya koymaktadır.

AYBİÖYÖ için temel bileşenler analizi kapsamında yapılan karşılaştırmalarda Kan’ın (2009) çalışmasında olduğu gibi kâğıt - kalem uygulamalarında metrik ölçeğin açıkladığı toplam varyans oranının Likert tipi ölçeğe göre daha yüksek olduğunu göstermiştir. Ayrıca internet üzerinden uygulanan metrik ölçek için de açıklanan varyans oranı kâğıt - kalem formatındaki Likert tipi ölçeğe göre daha yüksektir. Buna benzer şekilde AYKÖ için de metrik ölçek ile daha yüksek açıklanan varyans oranlarına ulaşılmıştır.

Gerich’e (2007) benzer şekilde kâğıt - kalem ortamındaki uygulamalarda metrik ölçek ile, internet ortamındaki uygulamalarda ise Likert ölçeği ile elde edilen güvenirlilik değerleri daha yüksektir. Ayrıca bu durumun, Gül’ün (2012) internet üzerinden daha yüksek güvenirlilik değerlerinin elde edildiği yönündeki bulgusuyla da örtüştüğü görülmektedir. AYBİÖYÖ için internet üzerinden Likert tipi ölçekle toplanan verilerin daha yüksek güvenirliliğe sahip olması ve AYKÖ için ise güvenirlilikler arasında manidar bir farkın olmaması Cook ve diğerlerinin (2001) metrik ölçeğin güvenirliliğinin daha yüksek olduğunu belirledikleri araştırmayla örtüşmemektedir. Bu durum diğer araştırmalarla bir bütün olarak değerlendirildiğinde çok da çelişkili görünmemektedir. Nitekim, karşılaştırılan bazı güvenirlilik değerleri için farkların manidar olmaması ve bu çalışmada olduğu gibi ayırt edicilik değerlerinin genel olarak metrik ölçek lehine olması Yıldız (1998), Koçyiğit (2002) ve Nartgün’ün (2002) çalışmalarına benzerlik göstermektedir.

Kâğıt - kalem ortamında uygulanan metrik ölçek ile Likert tipi ölçekten alınan puanlar arasındaki korelasyonların AYBİÖYÖ için .85; AYKÖ için ise .81 düzeyindedir. Bu durum, iki form arasındaki korelasyonun .74 ile .94 arasında değiştiğini belirten araştırmalar ile tutarlılık göstermektedir (Yıldız, 1998; Nartgün, 2002; Gerich, 2002; Atan ve Yavuz, 2010). İnternet üzerinden gerçekleştirilen uygulamalarda ise formlar arasındaki korelasyon değerlerinin .46 ve .77 gibi daha düşük değerler aldığı görülmüştür. Bu çalışmada cep telefonu ve bilgisayar kullanılarak verilen cevapların cihazın boyut ve diğer teknik özelliklerinin (ekran büyüklüğü, fare, kalem kullanımı vb.) hassasiyetinden etkilenmediği ve katılımcıların üniversite düzeyinde eğitim görmelerine bağlı olarak metrik ölçeğin internet uygulamasında herhangi bir zorluk yaşamadıkları varsayılmıştır. İnternet üzerinden yürütülen uygulamalarda formlar arasındaki korelasyonun düşük olması cevaplayıcıların yaşadığı bir zorlukla ilgili olabilir. Öğrencilerin cevaplama süreçlerine ilişkin herhangi bir inceleme gerçekleştirilmemiştir. Öte yandan, Caimmi ve diğerlerine (2017) benzer şekilde, önemli bir bölümü cep telefonundan elde edilen verilerin yüksek güvenirlilik ve geçerlik sağlayabildiği de ortadadır.

Farklı puanlayıcıların yaptığı ölçümler için maddeler arasındaki korelasyonların .90 ve üzerinde; toplam puanlar arasındaki korelasyonların ise mükemmel olması, Jensen ve diğerlerinin (1986) bahsettiği puanlama güvenirliliği sorunun önemli bir risk oluşturmadığını göstermiştir.

Kâğıt – kalem ve internet ortamında metrik ölçek ile elde edilen verilerin eşit aralıklara karşılık gelecek şekilde 2 – 20 arasında kategorilere ayrılması ile hesaplanan Cronbach α katsayılarındaki en büyük artışın Cicchetti, Showalter ve Tyrer (1985) tarafından yürütülen araştırmaya benzer şekilde iki kategoriden üç kategoriye geçildiğinde yaşandığı ve yedi kategoriden sonra önemli bir farkın kalmadığı belirlenmiştir. Buna göre Jacoby ve Matell'in (1971) üç kategorinin yeterli olacağı yönündeki önerisi desteklenmemektedir. Lissitz ve Green (1975), Lozano, Garcia-Cueto ve Muniz (2008), Preston ve Colman (2000) ve Wakita, Ueshima ve Noguchi (2012) gibi araştırmacıların yedi kategorinin kullanılması yönündeki önerileri daha makul görünmekte ve Likert tipi ölçeklerde yedi kategoriye yer verilmesi önerilmektedir.

Aun, Lam ve Collett (1986) ve Hui ve Triandis'in (1989) metrik ölçeklerin kullanımının kültüre göre farklılık yaratabileceği yönündeki ifadeleri de göz önüne alınırsa, yapılacak benzer araştırmalarda, bu araştırmada olduğu gibi yatay ve iki ucu etiketlenmiş metrik ölçeklerin kullanılması uygundur.

Araştırma kapsamında ulaşılan bulgular bir bütün olarak değerlendirildiğinde, göreceli olarak en zayıf psikometrik özelliklerin Likert tipi ölçeklerin kâğıt – kalem formatında uygulanmasıyla elde edildiği anlaşılmaktadır. Beş dereceli ölçeklerin metrik ölçekle değiştirilerek ya da internet üzerinden uygulanması daha iyi psikometrik özellikler sağlamaktadır. Uygulamaların kâğıt – kalem formatında yapıldığı durumlarda metrik ölçeklere, internet üzerinden yürütüldüğü durumlarda da Likert tipi ölçeklere yer verilmesi önerilebilir.

Araştırmanın farklı aşamalarında aynı öğrencilere ulaşmak mümkün olmamıştır. Dolayısıyla toplanan verilerde, birim yanıtlamama (unit nonresponse) biçiminde kayıp verilerin varlığından bahsedilebilir. Bu tür araştırmalarda kayıp veri tekniklerinin performansları karşılaştırılabilir. Benzer tasarımla yürütülen araştırmalarda toplanan veriler, çoklu özellik çoklu yöntem matrislerinin elde edilmesini sağlamaktadır. Kan'ın (2009) araştırmasında olduğu gibi, genellenebilirlik kuramı kapsamında kullanılan formların birer yüzey olarak ele alındığı araştırmalar planlanabilir. Bu araştırmada ölçeklerin psikometrik özellikleri temel bileşenler analizi, iç tutarlılık katsayıları, madde – toplam korelasyonları ve ölçüt bağıntılı geçerliği kapsamında değerlendirilmiştir. Benzer araştırmaların doğrulayıcı faktör analizi temelinde ölçme değişmezliğine yer verilecek şekilde tasarlanması düşünülebilir. Benzer şekilde, güvenilirliğe ilişkin incelemelerin eşdeğer formlar veya test – tekrar test güvenilirliği kapsamında ele alınması düşünülebilir.

Reips ve Funke (2008) tarafından hazırlanan yazılım internet ortamında metrik ölçek kullanımını sağlamakta fakat farklı tipte maddelerin uygulanması noktasında sınırlılıklar içermektedir. Atan ve Yavuz (2000) ve Abend ve diğerleri (2014) gibi araştırmacıların bu araştırmalara özel yazılımlar geliştirdiği, dolayısıyla bilgisayar ortamında uygulanan metrik ölçeklerin yaygınlaşmadığı düşünülmektedir. Son zamanlarda ise HTML5 form kontrollerinde bulunan range kontrolü ile metrik ölçeğin uygulanması kolaylaşmakta ve bu durumun ilerleyen zamanlarda bu ölçeğin yaygınlaşmasında olumlu etki yaratacağı düşünülmektedir. Bu çalışmada internet üzerinden metrik ölçeklerin uygulanmasına yönelik bir form geliştirilmiş ve bu form <http://kavramharitasiolustur.com/olcek/> adresinden yayınlanmıştır.

İnternet ortamında uygulanan metrik ölçekler ile bilgisayara yönelik tutum, bilgisayar okuryazarlığı gibi yapılar arasındaki ilişkiler incelenebilir. Bu gibi yapıların metrik ölçeklerin psikometrik özellikleri üzerindeki etkilerinin araştırılmasına ihtiyaç duyulmaktadır.

5. KAYNAKÇA

Abend, R., Dan, O., Maoz, K., Raz., S. & Bar-Haim, Y. (2014). Reliability, validity and sensitivity of a computerized visual analogue scale measuring state anxiety. *Journal of Behavior Therapy and Experimental Psychiatry*, 45(2014), 447-453.

Akat, K. (2012). *Ankilozan spondilitte hastalık aktivite ve fonksiyon ölçeklerinin Türkçe versiyonlarının farklı yanıt skalalarının performansının: nümerik derecelendirme skalası ve görsel analog skalasının karşılaştırılması* (Uzmanlık tezi). Dokuz Eylül Üniversitesi, Tıp Fakültesi, İç Hastalıkları Anabilim Dalı, İzmir.

Akbaş, U. & Tavşancıl, E. (2015). Farklı örneklem büyüklüklerinde ve kayıp veri örüntülerinde ölçeklerin psikometrik özelliklerinin kayıp veri baş etme teknikleri ile incelenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 6(1), 38-57.

Armstrong, R. L. (1987). The midpoint on a five-point Likert-type scale. *Perceptual and Motor Skills*, 64, 359-362.

Atan, M. & Yavuz, T. (2010). Davranışsal ölçümler için Likert-tipi ölçek / görsel analog ölçek karşılaştırması. 11. Ulusal Ekonometri ve İstatistik Sempozyumu, Sakarya Üniversitesi.

Athale, N., Sturley, A., Skoczen, S., Kavanaugh, A. & Lenert, L. (2004). A web-compatible instrument for measuring self-reported disease activity in arthritis. *The Journal of Rheumatology*, 31(2), 223-228.

Aun, C., Lam, Y. M. & Collet, B. (1986). Evaluation of the use of visual analogue scale in Chinese patients. *Pain*, 25(1986), 215-221.

- Ayaz, A. (2013). Evliliğe yönelik tutum ve mükemmeliyetçilik arasındaki ilişkinin incelenmesi. 10. Ulusal PDR Öğrencileri Kongresi. Dokuz Eylül Üniversitesi, 28-30 Haziran 2013 (sözlü bildiri)
- Baka, A., Figgou, L. & Triga, V. (2012). "Neither agree, nor disagree": a critical analysis of the middle answer category in voting advice applications. *International Journal of Electronic Governance*, 5(3/4), 244-263.
- Bandalos, D. L. & Enders, C. K. (1996). The effects of nonnormality and number of response categories on reliability. *Applied Measurement in Education*, 9(2), 151-160.
- Bardakçı, A., Haşiloğlu, S. B. & Balce, A. O. (2015). Ölçek orta noktası bitaraf cevaplar için mi yoksa bihaber cevaplar için mi kullanılmaktadır? *Pazarlama ve Pazarlama Araştırmaları Dergisi*, 16, 87-97.
- Bayat, B. (2014). Uygulamalı sosyal bilim araştırmalarında ölçme, ölçekler ve "Likert" ölçek kurma tekniği. *Gazi Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 16(3), 1-24.
- Baykul, Y. (2000). *Eğitimde ve psikolojide ölçme: Klasik Test Teorisi ve uygulama*. Ankara: ÖSYM Yayınları.
- Boer, A. G. E. M., Lanschot, J. J. B., Stalmeier, P. F. M., Sandick, J. W., Hulscher, J. B. F., Haes, J. C. J. M. & Sprangers, M. A. G. (2004). Is a single item visual analogue scale as valid, reliable and responsive as multi-item scales in measuring quality of life. *Quality of Life Research*, 13(2), 311-320.
- Bora Semiz, B. & Altunışık, R. (2016). Pazarlama araştırmalarında Likert tipi ölçeklerin özelliklerinin cevaplama tarzları üzerindeki etkilerinin incelenmesi. *Bartın Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 7(14), 577-598.
- Bosch, O. J., Revilla, M., DeCastellarnau, A. & Weber, W. (2018). Measurement reliability, validity, and quality of slider versus radio button scales in an online probability-based panel in Norway. *Social Science Computer Review*, 1-14.
- Büyüköztürk, Ş. (1997). Araştırmaya yönelik kaygı ölçeğinin geliştirilmesi. *Kuram ve Uygulamada Eğitim Yönetimi Dergisi*, 3(4), 453-464.
- Büyüköztürk, Ş. (2018). *Sosyal bilimler için veri analizi el kitabı: istatistik, araştırma deseni, SPSS uygulamaları ve yorum* (24. baskı). Ankara: Pegem Akademi.
- Caimmi, D., Baiz, N., Tanno, L. K., Demoly, P., Arnavielhe, S., Murray, R., Bedbrook, A., Bergmann, K. C., De Vries, G., Fokkens, W. J., Fonseca, J., Haahelta, T., Keil, T., Kuna, P., Mullol, J., Papadopoulos, N., Passalacqua, G., Samolinski, B., Tomazic, P. V., Valiulis, A., van Eerd, M., Wickman, M., Annesi-Maesano, I. & Bousquet, J. (2017). Validation of the MASK-rhinitis visual analogue scale on smartphone screens to assess allergic rhinitis control. *Clinical and Experimental Allergy: Journal of the British Society for Allergy and Clinical Immunology*, 47(12), 1526-1533.
- Carlsson, A. M. (1983). Assessment of chronic pain: I: aspects of the reliability and validity of visual analogue scale. *Pain*, 16(1983), 87-101.
- Cengiz Çakır, U. (2009). *Acil servise ağrı yakınması ile başvuran hastalarda hekimin ağrı algısını etkileyen faktörler ve görsel analog skala kullanımının değeri* (Uzmanlık tezi). Akdeniz Üniversitesi Tıp Fakültesi, Acil Tıp Anabilim Dalı, Antalya.
- Cengiz, K. B. (2013). *Transrektal ultrasonografi eşliğinde yapılan prostat biyopsilerinde uygulanan lokal anestezi yöntemlerinin visual analog skala (vas) ile karşılaştırılması* (Uzmanlık tezi). Cumhuriyet Üniversitesi, Tıp Fakültesi, Üroloji Anabilim Dalı, Sivas.
- Churchill, G. A. & Peter, J. P. (1984). Research desing effects on the reliability of rating scales: a meta-analysis. *Journal of Marketing Research*, 21(4), 360-375.
- Cicchetti, D. V., Showalter, D. & Tyrer, P. J. (1985). The effect of number of rating scale categories on levels of interrater reliability: a monte carlo investigation. *Applied Psychological Measurement*, 9(1), 31-36.
- Coelho, P. S. & Esteves, S. P. (2007). The choice between a five-point and a ten-point scale in the framework of customer satisfaction measurement. *International Journal of Market Research*, 49(3), 313-339.
- Cook, C., Heath, F., Thompson, R. L. & Thompson, B. (2001). Score reliability in web- or internet-based surveys: unnumbered graphic rating scales versus Likert-type scales. *Educational and Psychological Measurement*, 61(4), 697-706.

- Coons, S. J., Gwaltney, C. J., Hays, R. D., Lundy, J. J., Sloan, J. A., Revicki, D. A., Lenderking, W. R., Cella, D. & Basch, E. (2009). Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO good research practices task force report. *Value in Health*, 12(4), 419-429.
- Cox III, E. P. (1980). The optimal number of responses for a scale: a review. *Journal of Marketing Research*, 17(4), 407-422.
- Cummins, R. A. & Gullone, E. (2000). Why we should use 5-point Likert scales: the case for subjective quality of measurement. Proceedings, Second International Conference on Quality of Life in Cities (pp. 74-93). Singapore: National University of Singapore.
- Dawes, J. (2008). Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales. *International Journal of Market Research*, 50(1), 61-77.
- Delgado, D. A., Lambert, B. S., Boutris, N., McCulloch, P. C., Robbins, A. B., Moreno, M. R. & Harris, J. D. (2018). Validation of digital visual analog scale pain scoring with a traditional paper-based visual analogue scale in adults. *Journal of the American Academy of Orthopaedic Surgeons*, 2(3), 1-6.
- Demir, Ö. O. (2014). Üniversite öğrencilerinin akış deneyimleri ve kişilik özellikleri arasındaki ilişkinin incelenmesi. 11. Ulusal PDR Öğrenci Kongresi. Hacettepe Üniversitesi, Ankara. Sözlü bildiri.
- DiStefano, C. (2002). The impact of categorization with confirmatory factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(3), 327-346.
- Dixon, J. S. & Bird, H. A. (1981). Reproducibility along a 10 cm vertical visual analogue scale. *Annals of Rheumatic Diseases*, 40, 87-89.
- Dixon, P. N., Bobo, M. & Stevick, R. A. (1984). Response differences and preferences for all-category-defined and end-defined Likert formats. *Educational and Psychological Measurement*, 44, 61-66.
- Erkuş, A. (2016). *Psikolojide ölçme ve ölçek geliştirme-I: temel kavramlar ve işlemler* (3. baskı). Ankara: Pegem Akademi.
- Erkuş, A., Sanlı, N., Bağlı, M. T. & Güven, K. (2000). Öğretmenliğe ilişkin tutum ölçeği geliştirilmesi. *Eğitim ve Bilim*, 25(116), 27-33.
- Finn, R. H. (1972). Effects of some variations in rating scale characteristics on the means and reliabilities of ratings. *Educational and Psychological Measurement*, 32, 255-265.
- Garland, R. (1991). The mid-point on a rating scale: is it desirable? *Marketing Bulletin*, 2, 66-70.
- Gerich, J. (2007). Visual analogue scales for more independent measurement in self-administered questionnaires. *Behavior Research Methods*, 39, 985-992.
- Gök, B., Atalay Kabasakal, K. & Çetin, S. (2015). Araştırmanın yöntem bölümüne ilişkin öz yeterlik ölçeği geliştirilmesi. *İlköğretim Online*, 14(2), 438-448.
- Gül, E. (2012). Online değerlendirme güvenilir midir? *e-Journal of New World Sciences Academy*, 7(1), 281-287.
- Harland, N. J., Dawkin, M. J. & Martin, D. (2015). Relative utility of a visual analogue scale vs a six-point Likert scale in the measurement of global subject outcome in patients with low back pain receiving physiotherapy. *Physiotherapy*, 101, 50-54.
- Harpe, S. E. (2015). How to analyze Likert and other rating scale data. *Currents in Pharmacy Teaching and Learning*, 7, 836-850.
- Hawker, G. A., Mian, S., Kendzerska, T. & French, M. (2011). Measures of adult pain. *Arthritis Care & Research*, 63(11), 240-252.
- Hughes, G. D. (1969). Some confounding effects of forced-choice scales. *Journal of Marketing Research*, 6(2), 223-226.
- Hui, C. H. & Triandis, H. C. (1989). Effects of culture and response format on extreme response style. *Journal of Cross-Cultural Psychology*, 20(3), 296-309.
- Huskisson, E. C. (1974). Measurement of pain. *The Lancet*, 9, 1974.

- İnal, H., Yılmaz Koğar, E., Demirdüzen, E. & Gelbal, S. (2017). Cronbach's coefficient alpha: a meta-analysis study. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 32(1), 18-32.
- Jacoby, J. & Matell, M. S. (1971). Three-point Likert scales are good enough. *Journal of Marketing Research*, 8(4), 495-500.
- Jensen, M. P., Caroly, P. & Braver, S. (1986). The measurement of clinical pain intensity: a comparison of six methods. *Pain*, 27(1986), 117-126.
- Jöreskog, K. (1999). How large can a standardized coefficient be? Unpublished document.
- Kan, A. (2009). Effect of scale response format on psychometric properties in teaching self-efficacy. *Eurasian Journal of Educational Research*, 34, 215-228.
- Koçyiğit, E. (2002). *Aynı tutumu ölçmeye yönelik olarak hazırlanmış Likert tipi ölçekle iki farklı metrik ölçeğin psikometrik özelliklerinin karşılaştırılması* (Yayınlanmamış yüksek lisans tezi). Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Laerhoven, H., van der Zaag-Loonen, H. J. & Derkx, B. H. F. (2004). A comparison of Likert scale and visual analogue scales as response options in children's questionnaires. *Acta Pædiatr*, 93, 830-835.
- Leung, S. (2011). A comparison of psychometric properties and normality in 4-, 5-, 6-, and 11-point Likert scales. *Journal of Social Service Research*, 37, 412-421.
- Lissitz, R. W. & Green, S. B. (1975). Effect of number of scale points on reliability: a monte carlo approach. *Journal of Applied Psychology*, 60(1), 10-13.
- Lozano, L. M., Garcia-Cueto, E. & Muniz, J. (2008). Effect of number of response categories on the reliability and validity of rating scales. *Methodology*, 4(2), 73-79.
- Maharajan, M. K., Rajiah, K., Khoo, S. P., Chellappan, D. K., Alwis, R. D., Chui, H. C., Tan, L. L., Tan, Y. N. & Lau, S. Y. (2017). Attitudes and readiness of students healthcare professions towards interprofessional learning. *PLoS ONE*, 12(1): e0168863.
- Mattel, M. S. & Jacoby, J. (1972). Is there an optimal number of alternatives for Likert-scale items? Effects of testing time and scale properties. *Journal of Applied Psychology*, 56(6), 506-509.
- McMordie, W. R. (1979). Improving measurement of death anxiety. *Psychological Reports*, 44, 975-980.
- Miller, S. D., Duncan, B. L., Brown, J., Sparks, J. A. & Claud, D. A. (2003). The outcome rating scale: a preliminary study of the reliability, validity, and feasibility of a brief visual analogue scale. *Journal of Brief Therapy*, 2(2), 91-100.
- Myles, P., Troedel, S. Boquest, M. & Reeves, M. (1999). The pain visual analogue scale: Is it linear or nonlinear? *Anesthesia and Analgesia*, 89, 1517-1520.
- Nartgün, Z. (2002). *Aynı tutumu ölçmeye yönelik Likert tipi ölçek ile metrik ölçeğin madde ve ölçek özelliklerinin Klasik Test Kuramı ve Örtük Özellikler Kuramı'na göre incelenmesi* (Yayınlanmamış doktora tezi). Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education: Theory and Practice*, 15, 625-632.
- Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill, Inc.
- Ogon, M., Krismer, M., Söllner, W., Kantner-Rumplmair, W. & Lampe, A. (1996). Chronic low back pain measurement with visual analogue scale in different settings. *Pain*, 64(1996), 425-428.
- Osanmaz, B., Bilgi, E., İnce, M. & Tanju Aslışen, H. E. (2018). Aile içi şiddetin çocuk üzerindeki etkisi. 13. Okul Öncesi Eğitimi Öğrenci Kongresi, Kocatepe Üniversitesi, Afyonkocatepe Üniversitesi, 26-28 Nisan 2018 (poster bildiri).
- Pearse, N. (2011). Deciding on the scale granularity of response categories of Likert type scales: the case of a 21-point scale. *The Electronic Journal of Business Research Methods*, 9(2), 159-171.
- Pedhazur, E. J. & Schmelkin, L. (1991). *Measurement, design, and analysis: an integrated approach*. New Jersey: Lawrence Erlbaum Associates, Inc.

- Peterson, R. A. (1994). A meta-analysis of Cronbach's coefficient alpha. *Journal of Consumer Research*, 21(2), 381-391.
- Raaijmakers, Q. A. W., Van Hoof, A., Hart, H., Verbogt, T. F. M. A. & Vollebergh, W. A. M. (2000). Adolescents' midpoint responses on Likert-type scale items: neutral or missing values. *International Journal of Public Opinion Research*, 12(2), 208-216.
- Preston, C. C. & Colman, A. M. (2000). Optimal response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104, 1-15.
- Price, D. D., McGrath, P. A., Rafi, A. & Buckingham, B. (1983). The validation of visual analogue scale as ratio scale measures for chronic and experimental pain. *Pain*, 17(1983), 45-56.
- Reips, U. & Funke, F. (2008). Interval-level measurement with visual analogue scales in internet-based research: VAS Generator. *Behavior Research Methods*, 40(3), 699-704.
- Saracaloğlu, A. S. (2008). Lisansüstü öğrencilerin akademik güdülenme düzeyleri, araştırma kaygıları ve tutumları ile araştırma yeterlikleri arasındaki ilişki. *Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi*, 5(2), 179-208.
- Snow, S. & Kirwan, J. R. (1988). Visual analogue scales: a source of error. *Annals of Rheumatic Diseases*, 47, 526.
- Tabachnick, B. G. & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). New Jersey: Pearson.
- Tavşancıl, E. (2002). *Tutumların ölçülmesi ve SPSS ile veri analizi*. Ankara: Nobel Yayın Dağıtım.
- Tezbaşaran, A. (1996). *Likert tipi ölçek geliştirme kılavuzu*. Ankara: Psikologlar Derneği Yayınları.
- Turan, İ., Şimşek, Ü. & Aslan, H. (2015). Eğitim araştırmalarında Likert ölçeği ve Likert-tipi soruların kullanımı ve analizi. *Sakarya Üniversitesi Eğitim Fakültesi Dergisi*, 30, 186-203.
- Uyumaz, G. (2013). *Likert tipi ölçeklerde madde düzeni ve derecelendirme farklılıklarının psikometrik özellikler ve yanıtlayıcı tutumları açısından karşılaştırılması*. Yayımlanmamış yüksek lisans tezi, Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
- Wakita, T., Ueshima, N. & Noguchi, H. (2012). Psychological distance between categories in the Likert scale: comparing different numbers of options. *Educational and Psychological Measurement*, 72(4), 533-546.
- Watkins, M. W. (2000). *Monte carlo PCA for parallel analysis* [computer software]. State College, PA: Ed & Psych Associate.
- Weng, L. (2004). Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educational and Psychological Measurement*, 64(6), 956-972.
- Yıldırım, S. (2011). Öz-yeterlik, içe dönük motivasyon, kaygı ve matematik başarısı: Türkiye, Japonya ve Finlandiya'dan bulgular. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi*, 5(1), 277-291.
- Yıldız, Y. (1998). *Aynı tutumu ölçmeye yönelik Likert tipi ölçek ile metrik ölçeğin psikometrik özelliklerinin karşılaştırılması* (Yayımlanmamış doktora tezi). Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- YÖK (Yüksek Öğretim Kurulu), 2018. *Rehberlik ve Psikolojik Danışmanlık Lisans Programı*. [http://www.yok.gov.tr/documents/10279/41805112/Rehberlik ve Psikolojik Danismanlik Lisans Programi.pdf](http://www.yok.gov.tr/documents/10279/41805112/Rehberlik+ve+Psikolojik+Danismanlik+Lisans+Programi.pdf) adresinden 1 Ağustos 2018 tarihinde erişildi.
- Zumbo, B. D., Gadermann, A. M. & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for Likert type scales. *Journal of Modern Applied Statistical Methods*, 6(1), 21-29.

6. EXTENDED ABSTRACT

Likert type scales are frequently used when measuring the variables subject to education and psychology. In addition to its advantages and widespread use, there are ongoing discussions on issues such as the presence of obtained data in ordinal scale, whether midpoints can be reached in categories or not, how categories should be labeled, whether all categories should be labeled or not, and the number of categories. Metric scales can be applied in different formats and they constitute a significant alternative to Likert type scales.

It is also known that when metric scales are used, there are certain limitations regarding the respondent's position towards the related variable, respondent's matching and marking of the point on scale and the conversion of measurements into data through the use of a schedule. First of these limitations also apply for Likert type scales. Second limitation can be overcome through applications realized via computer. Despite these advantages, it is observed that the metric scales which are frequently used in measuring the variables such as pain, fatigue, and sensitivity in the field of health are used relatively less in the field of education. The most common use of metric scales which can be applied in different forms (horizontal / vertical, simple, numerical, etc.) is the form in which two ends are labeled and the mid-point is specified. Studies indicate that the most suitable metric scale is a 10 cm long, horizontal metric scale marked on middle and labeled on both ends.

This study is designed in relational type and the psychometric traits of two different scales are examined through the five category Likert type scale and metric scale applied via paper pencil tests and web-based applications. In this context, the relationship between the Self-Efficacy Scale about the Method Sections of the Research Papers – SESMSR, which is developed to apply on graduate students, with The Research Anxiety Scale – TRAS within the scope of psychometric traits of undergraduate students and criterion validity.

A total of 391 students enrolled in the Faculties of Education of two different universities in the spring semester of the 2017 - 2018 academic year participated in the study. The psychometric traits of the scales were examined separately with the data obtained by applying them in two different conditions and two different formats. When the descriptive statistics of the total scores from the scales were examined, it was seen that the distributions for all applications were considered normal.

Analysis of the key components for SESMSR indicated that the two-factor structure of the original form emerged with the exception of one item falling within the other factor. Reliability and item - total correlations were high for all data sets. The total scores obtained from the SESMSR and overall factors were in favor of the graduate students when compared with undergraduate students; and for those who attended the scientific research methods course when compared with those who did not. The evaluation of the findings obtained for SESMSR as a whole indicated that the scale had psychometric traits suitable for use in the studies conducted on undergraduate level students. Analysis of the key components for TRAS indicated that data obtained from different applications revealed the single factor structure of the original form. Moreover, the reliability coefficients and item - total correlations of the scale were high.

The comparison of the results obtained from two different data sets were examined and it was found out that the lowest extracted variance rate was obtained from the Likert type scale applied as paper and pencil test. The highest explanatory variance rate for SESMSR was obtained from a Likert - type scale via web-based application, while the highest variance rate for TRAS was obtained from a metric scale applied as paper and pencil test. The reliability of the scales was examined by using the Cronbach α coefficient and the reliability of the equivalent halves and it was seen that the lowest values were obtained from the Likert type scale applied as paper and pencil test. The significance of the difference between the reliability coefficients was examined by using the Fisher z test and it was determined that more reliable measurements were obtained from the metric scale applied as paper and pencil test when compared to the Likert type scale.

The correlations obtained from the measurements of all applications were observed and it was seen that there were medium/high level relations between the different applications of same scale. The highest correlations for SESMSR total scores were measured by the metric scale applied in paper and pencil and the highest correlations for the TRAS total scores were measured by the Likert type scale applied in paper and pencil.

As the findings of research were evaluated as a whole, it was understood that the lowest psychometric traits were obtained from a five - point Likert - type scale applied as paper and pencil test. Researchers are recommended to use metric scale instead of five-point Likert-type scale, and if they decide to use Likert-type scale then they should include seven categories or use the web based application.

New research is needed on the use of metric scales in the fields of education and psychology. Within this scope, similar research may be scheduled in which the number of items, factors or degrees in scales are manipulated. The studies can be planned in a way that multi-trait multi-method matrices are examined or the studies can be performed based on generalizability theory through web-based and paper and pencil test applications. Moreover, in this study, it was assumed that university students did not have any problems in terms of web-based applications and therefore no examinations were conducted in relation with this. New research may include structures such as attitude toward the Internet and computer literacy.