Bolu Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi (BAİBÜEFD)

Bolu Abant Izzet Baysal University Journal of Faculty of Education

2025, 25(3), 1567 – 1585. DOI: 10.17240/aibuefd.2025..-1563316

Eğitimde Yapay Zekâ Kullanımına İlişkin Geliştirilen ve Uyarlanan Ölçeklerin İncelenmesi

Adequacy of the Scales Developed for the Use of Artificial Intelligence in Education

Nursel YILMAZ[1], **Özgen KORKMAZ**[2]

Geliş Tarihi (*Received*): 08.10.2024 **Kabul Tarihi (*Accepted*):** 10.06.2025 **Yayın Tarihi (*Published*):** 15.09.2025

Öz: Bu çalışmada, eğitimde yapay zekâ kullanımına yönelik ölçeklerin geçerlik ve güvenirlik açısından incelenmesi amaçlanmıştır. Çalışmada yer alan 24 ölçek, ölçeklerin türleri, geliştirilme amaçları, kullanılan geçerlik ve güvenirlik analizleri ile çalışma grupları dikkate alınarak analiz edilmiştir. Ölçeklerin üçte biri farklı dillerden Türk kültürüne uyarlanmış, üçte ikisi ise bağımsız olarak geliştirilmiştir. Ölçeklerin büyük kısmı yapay zekâ okuryazarlığı ve tutum konularına odaklanmaktadır. Çalışma grubu olarak genellikle üniversite öğrencileri tercih edilmiş olup öğretmenler ve K12 öğrencilerine yönelik ölçeklerin sayısı sınırlıdır. Araştırmada eğitimde yapay zekâya dair dört tema oluşturulmuş; ölçeklerin amacı, örneklem, geçerlik ve güvenirlik analizleri belirlenmiştir. Geçerlik analizlerinde en yaygın kullanılan yöntem yapı geçerliği olup, özellikle açımlayıcı (AFA) ve doğrulayıcı faktör analizlerinin (DFA) birlikte kullanıldığı görülmektedir. Ancak, bu analizlerde farklı örneklemlerle çalışma yapılmadığı ve ayırt edicilik özelliğinin yeterince araştırılmadığı tespit edilmiştir. Güvenirlik analizlerinde, ölçeklerin büyük çoğunluğunda Cronbach alfa katsayısı hesaplanmış ve yüksek güvenirlik oranları elde edilmiştir. Ancak test tekrar test yöntemine daha az başvurulmuştur. Sonuç olarak, eğitimde yapay zekâ kullanımına yönelik geliştirilen ölçekler genel olarak geçerlik ve güvenirlik açısından yeterli görülmektedir. Ancak, çalışma gruplarının çeşitlendirilmesi ve eksik analizlerin tamamlanması önerilmektedir. Özellikle K12 düzeyindeki öğrenciler ve öğretmenler için ölçek geliştirme çalışmalarına ihtiyaç duyulmaktadır.

Anahtar Kelimeler: Eğitimde yapay zekâ kullanım, Ölçek geliştirme, Betimsel analiz.

&

Abstract: The study aims to examine the scales for the use of artificial intelligence in education in terms of validity and reliability. The 24 scales in the study were analyzed by considering the types of scales, the purposes of their development, the validity and reliability analyses used, and the study groups. One third of the scales were adapted from different languages into Turkish and two thirds were developed independently. Most of the scales focus on artificial intelligence literacy and attitudes. University students were generally preferred as the study group, and the number of scales for teachers and K12 students is limited. In the study, four themes regarding artificial intelligence in education were created; the purpose, sampling, validity and reliability analyses of the scales were determined. The most used method in validity analyses is construct validity, and especially exploratory (EFA) and confirmatory factor analyses (CFA) are used together. However, in these analyses, it was found that studies with different samples were not conducted, and discriminant validity was not paid enough attention. In reliability analyses, Cronbach's alpha coefficient was calculated for the majority of the scales and high reliability rates were obtained. However, the test-retest method was used less frequently. As a result, although the scales developed for the use of artificial intelligence in education are generally considered sufficient in terms of validity and reliability, it is recommended that the study groups be diversified, and the missing analyzes be completed. There is a need for scale development studies especially for K12 level students and teachers.

Keywords: Use of artificial intelligence in education, Ccale development, Content analysis.

Atıf/Cite as: Yılmaz, N., Korkmaz, Ö. (2025). Eğitimde yapay zekâ kullanımına ilişkin geliştirilen ve uyarlanan ölçeklerin incelenmesi. *Bolu Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 25(3), 1567-1585, DOI: 10.17240/aibuefd.2025..-1563316

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

[1] Nursel Yılmaz, Amasya Üniversitesi, Bilgisayar ve Öğretim Teknolojileri Eğitimi Anabilim Dalı, e-mail: nyilmazmat@gmail.com, ORCID: https://orcid.org/0000-0001-6708-0656

[2] Sorumlu Yazar: Prof. Dr. Özgen Korkmaz, Bandırma Onyedi Eylül Üniversitesi, Bilgisayar Mühendisliği Bölüm, e-mail: ozgenkorkmaz@bandirma.edu.tr, ORCID: https://orcid.org/0000-0003-4359-5692

1. GİRİŞ

Toplumsal dönüşüm, yalnızca endüstriyel ve teknolojik gelişmelerle sağlanamaz; eğitim ve sağlık alanlarında da sistematik değişimler gereklidir (Öztemel, 2018). Son 25 yılda Yapay Zekâ (YZ) tabanlı eğitim teknolojilerinde önemli ilerlemeler kaydedilmiş olsa da bu araçların eğitim ortamlarında geniş çapta benimsenmesi sınırlı kalmıştır (Roll & Wylie, 2016). Medyada YZ'ye yönelik olumsuz söylemler, bu teknolojilerin benimsenmesini olumsuz etkileyebilir (Frey & Osborne, 2013). Ancak, Xie ve Wang'ın (2024) çalışması, YZ tabanlı etkileşimli öğrenme ortamlarının öğrencilerin bilişsel yetilerini olumsuz etkilemediğini göstermektedir. Bununla birlikte, okullarda YZ tabanlı eğitim teknolojilerine karşı direnç gözlemlenmektedir (Nazaretsky vd., 2022). YZ'nin meslekler üzerindeki etkisi kaçınılmaz olup, bireylerin gelecekteki mesleklerde ihtiyaç duyulacak becerilerle donatılması önem taşımaktadır. Bu bağlamda, yapay zekâ okuryazarlığı, bireylerin YZ teknolojilerini eleştirel değerlendirme ve etkili kullanma yetkinliğini ifade etmektedir (Yavuz Aksakal & Ülgen, 2021; Long & Magerko, 2020).

Eğitimde yapay zekâya (YZ) yönelik ilgi giderek artmış, ancak derin öğrenme teknolojileri üzerine yapılan çalışmalar sınırlı kalmış, daha çok doğal dil işleme teknolojilerine odaklanılmıştır (Chen, Xie, Zou & Hwang, 2020). Araştırmalar genellikle özel eğitimde akıllı ders sistemleri, dil eğitimi için doğal dil işleme, yapay zekâ eğitimi için eğitici robotlar, performans tahmini için eğitsel veri madenciliği ve kişiselleştirilmiş öğrenme için öneri sistemleri gibi konulara yönelmiştir (Chen vd., 2022). YZ tabanlı uyarlanabilir öğrenme sistemleri, her öğrencinin yetenek ve tercihine uygun eğitim materyalleri sunarak öğrenme sürecini bireyselleştirmektedir (Guan vd., 2020). YZ'nin eğitimde kullanımı, gelişmiş verimlilik, küresel öğrenme, kişiselleştirilmiş öğretim ve eğitim yönetiminde etkinlik sağlamaktadır (Timms, 2016). Eğitim sistemleri, teknolojik gelişmelere uyum sağlamak zorundadır ve toplumların dijital dönüşüm yol haritalarını belirlemesi kaçınılmazdır (Çoban & Uzun, 2022; Öztemel, 2018). YZ, makine öğrenimi ve uyarlanabilirlik sayesinde müfredat ve içeriği öğrencilerin ihtiyaçlarına göre özelleştirebilmekte, böylece öğrenme deneyimini iyileştirmektedir (Chen vd., 2020). Ancak, YZ tabanlı bir eğitim ekosistemi oluşturmak için kamu politikalarının uluslararası ve ulusal düzeyde iş birliği içinde olması, öğretmenlerin dijital becerilerini geliştirmesi ve YZ geliştiricilerinin eğitim süreçlerine uyum sağlaması gerekmektedir (Pedro vd., 2019). Öğretmenler YZ platformlarını kullanarak öğrencilerin çalışmalarını daha etkili değerlendirebilir ve öğretim kalitesini artırabilirler (Chen vd., 2020). Ayrıca, YZ'nin öğretim süreçlerine entegrasyonu, robotların öğretmen asistanı olarak kullanılmasıyla öğrencilere okuma ve telaffuz gibi görevlerde destek sağlamasına olanak tanımaktadır (Timms, 2016). YZ teknolojilerinin öğretim süreçlerinde daha yaygın kullanımı, eğitmenlerin tekrarlayan işlerden kurtulmasını ve öğrencilere daha hızlı geri bildirim sunmasını sağlayarak uyarlanabilir ve kişiselleştirilmiş öğretimi geliştirmektedir (Chan & Zary, 2019). YZ, eğitim kurumlarında idari görevleri otomatikleştirme konusunda da önemli bir potansiyele sahiptir (Timms, 2016). Eğitim sisteminin gelecekteki gelişimi, YZ'nin ilerlemesiyle doğrudan ilişkili olup, akıllı makinelerin bilgi işlem kapasitelerinin artışıyla daha da ileri taşınacaktır (Chen vd., 2020).

YZ teknolojilerinin eğitim alanında daha etkili ve yaygın bir şekilde kullanılabilmesi, yalnızca teknik yeterliliklerle değil, aynı zamanda eğitim paydaşlarının YZ'ye yönelik olumlu algılar, tutumlar ve yeterliliklerle donatılmasını da gerektirmektedir. Özellikle öğretmenler, öğrenciler ve eğitim yöneticilerinin YZ teknolojilerine yönelik olumlu bakış açılarına sahip olmaları, bu teknolojilerin benimsenmesi ve etkili bir şekilde uygulanması açısından kritik bir faktördür. Bu bağlamda, bireylerin YZ okuryazarlık düzeyleri, YZ'ye duydukları güven, YZ tabanlı eğitim sistemlerine yönelik öz yeterlilikleri ve YZ'nin eğitimdeki rolüne ilişkin tutumları, teknolojinin sürdürülebilir bir şekilde entegrasyonunda belirleyici olmaktadır. Bu tür değişkenler, psikolojik ve bilişsel faktörler olarak sınıflandırılabilir ve eğitimde YZ'ye yönelik algıları ölçmek amacıyla farklı ölçeklerle incelenmektedir. Alanyazında, YZ farkındalık düzeyini (Ferikoğlu & Akgün, 2022), YZ okuryazarlığını (Çelebi vd., 2023; Hwang vd., 2023; Laupichler vd., 2023; Karaoğlan Yılmaz & Yılmaz, 2023; Ng vd., 2023; Polatgil & Güler, 2023; Wang vd., 2023), YZ tabanlı eğitim teknolojilerine güveni (Nazaretsky vd., 2022), YZ öz yeterliliğini (Wang & Chuang, 2024; Morales-García vd., 2024), YZ öğreniminin karşılaştırmalı öz değerlendirmesini (Laupichler vd.,

2023) ve YZ kullanımına ilişkin görüşleri belirlemeye yönelik ölçekler geliştirilmiştir (Dülger & Köklü, 2023). Bunun yanı sıra, yapay zekâya hazırlık (Karaca vd., 2021), YZ'nin kabulü (Yılmaz vd., 2023), YZ kaygısı (Akkaya vd., 2021; Terzi, 2020; Wang & Wang, 2019), YZ eğitimi algısı (Chan & Zhou, 2023; Cukurova vd., 2020), YZ'ye yönelik tutum (Kaya vd., 2024; Schepman & Rodway, 2023; Suh & Ahn, 2022), YZ etiği (Jang vd., 2022), YZ bağımlılığı (Morales-García vd., 2024) ve meta-eğitim inançlarını (Erol vd., 2023) ölçmeye yönelik ölçeklerin alanyazında yer aldığı görülmektedir. Bu ölçekler, bireylerin YZ'ye yönelik bilişsel ve duyuşsal tepkilerini analiz etmeye yardımcı olmakta ve YZ'nin eğitim sistemlerine entegrasyonunu şekillendiren önemli değişkenleri belirlemektedir.

Psikolojik bir değişkenin ölçülmesi için geliştirilen ölçekler, var olan bir ölçeğin eksikliklerinin görülmesi veya yeni bir değişkenin incelenmesi için geliştirilir (Erkuş, 2019). Kişilerin ilgi, tutum ve güdülenmişlik gibi duyuşsal alan durumları ilgi envanterleri ve bireylerin bir ya da birçok boyutta kendi tutumlarının rapor ettikleri en popüler araç olan tutum ölçekleri ile ölçülebilir (Demirel, 2020). Psikolojik ölçme aracı geliştirmede akla öncelikle Likert tipi ölçekler gelmektedir, başlangıçta tutum ölçmek için geliştirilen bu teknik günümüzde pek çok psikolojik değişkeni ölçmek için kullanılmaktadır (Acar Güvendir & Özer Özkan, 2022). Bir ölçme aracında geçerlik, güvenirlik gibi niteliklerin bulunması arzulanır (Tan, 2020). Bir ölçme aracının geçerli olabilmesi için öncelikle güvenilir olması gerekir, bu iki nitelik ölçek geliştirme sürecinin ta kendisidir (Erkuş, 2019). Ölçme aracının hatasız, kararlı, tutarlı ve duyarlı ölçüm yapma derecesi güvenirlik; amaçladığı özelliği başka faktörlerle karıştırmadan ölçme derecesi ise geçerlik olarak tanımlanır (Başol, 2019).

Ölçeklerin geçerlik ve güvenirliğini değerlendirmek için çeşitli analizler yapılır (Büyüköztürk, 2023; DeVellis, 2017; Hair vd., 2014; Koğar, 2021; Nunnally & Bernstein, 1994;). Geçerlik analizleri: yapı geçerliği, kapsam geçerliği, ayırt edicilik, amaca hizmet etme düzey ve kriter geçerliği şeklinde yapılır. Ölçeğin yapısının doğruluğu, açımlayıcı faktör analizi ve doğrulayıcı faktör analizi gibi tekniklerle değerlendirilirken; ölçeğin ölçülmek istenen kavramı tüm yönleriyle kapsadığını belirlemek için kapsam geçerliğine bakılır. Madde ayırt ediciliği, bir test maddesinin düşük ve yüksek yetenek düzeyine sahip bireyleri ayırt etme gücünü ifade eder ve Madde Tepki Kuramı (MTK) çerçevesinde ele alınarak daha hassas ölçümler yapılmasına olanak tanır (Lord, 2012; Hambleton vd., 1991). Geleneksel ayırt edicilik indeksleri genellikle tek boyutlu yapıları temel alırken, iki boyutlu testlerde bu değerlerin ne kadar gerçekçi olduğu sorgulanmış ve bu yapılar için, maddelerin farklı boyutlardaki ayırt ediciliğini ayrı ayrı hesaplayan alternatif bir indeks önerilmiştir (Kılıç & Uysal, 2022). Likert tipi ölçeklerde madde ayırt ediciliğinin hesaplanmasına yönelik yeni bir yaklaşım ise MTK tabanlı çok kategorili modellerin kullanımını teşvik ederek, maddelerin farklı tepki kategorilerinde ayırt edicilik seviyelerini daha hassas bir şekilde belirlemeyi amaçlamaktadır (Çelen & Aybek, 2022). Bununla birlikte ölçek maddelerinin tutarlılığını ve kararlılığı belirleyen güvenirlik analizleri için iç tutarlık ve test-tekrar test teknikleri kullanılır. İç tutarlık güvenirliği için Cronbach alfa, eş yarılar arasındaki korelasyon gibi değerler ölçülürken; ölçeğin aynı bireylere farklı zamanlarda uygulanması ve sonuçların tutarlılığının aralarındaki korelasyon ile belirlenmesi için test -tekrar test tekniği kullanılır. Ancak ölçek geliştirme süreci ve metodolojileri ile ilgili literatürde bir birlikteliğin olmadığını söylemek mümkündür. Bunun sonucu olarak da her ne kadar yayınlanmış da olsa literatürde bulunan ölçeklerin geçerlik ve güvenirlik açısından uygun olup olmadığı ile ilgili soru işaretleri söz konusudur. Bu çalışmada eğitimde YZ kullanımına yönelik literatürde bulunan ölçeklerin geçerlik ve güvenirliğini belirlemek için yapılan çalışmaların yeterliliği ele alınmıştır.

1.1. Araştırmanın amacı

Çalışmanın amacı, eğitimde yapay zekâ kullanımına ilişkin geliştirilmiş ve uyarlanmış ölçeklerin geçerlik ve güvenirlik analizlerini incelemektir. Bu çerçevede aşağıdaki incelemeler gerçekleştirilmiştir:

1- Eğitimde yapay zekâ kullanımına yönelik ölçeklerin amaçlarını belirlemek.
2- Geliştirilen ölçeklerin Türk kültürüne uyarlanması veya ölçek geliştirme süreçlerini incelemek.
3- Örneklem grubu ve büyüklüğünü betimlemek.
4- Geçerlik analizlerini (yüzey, kapsam, kriter, yapı geçerliği) incelemek.
5- Güvenirlik analizlerini (iç tutarlılık ve test tekrar test) incelemek.

1.2. Araştırmanın önemi

Eğitimde yapay zekâ (YZ) teknolojilerinin kullanımının giderek artması, bu alanda geliştirilen ölçeklerin geçerlik ve güvenirliğinin incelenmesini önemli hâle getirmektedir. Eğitimde yapay zekâ, öğretim süreçlerini kişiselleştirmek (Luckin, 2017), öğrenci başarısını artırmak (Holmes vd., 2019) ve öğretmenlere destek sağlamak (Zawacki-Richter vd., 2019) gibi birçok avantaj sunmaktadır. Ancak, bu teknolojilerin etkin bir şekilde kullanılabilmesi için doğru ölçme araçlarına ihtiyaç duyulmaktadır.

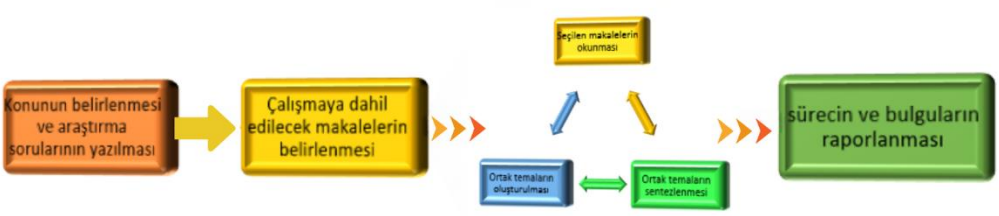
Bu çalışmada, eğitimde yapay zekâ kullanımına yönelik ölçeklerin geçerlik ve güvenirlik analizlerinin detaylı olarak incelenmesi, eğitim araştırmalarında sağlam temellere dayanan verilerin elde edilmesine katkı sunacaktır. Literatürde, bir ölçeğin psikometrik açıdan yeterli olması için örneklem büyüklüğünün uygun olması (Gül & Sözbilir, 2015), yapı geçerliğinin detaylı olarak analiz edilmesi (Cabrera-Nguyen, 2010) ve güvenirliğinin yüksek olması gerektiği vurgulanmaktadır (Ergene, 2020). Bu kapsamda, çalışmamız eğitimde yapay zekâ kullanımına ilişkin mevcut ölçeklerin bilimsel standartlara uygunluğunu değerlendirerek, araştırmacılar için önemli bir kaynak oluşturmayı amaçlamaktadır.

Ayrıca, ölçeklerin Türk kültürüne uyarlanması ve yeni ölçeklerin geliştirilmesi, eğitim araştırmalarında yerel bağlamın dikkate alınmasını sağlayacaktır (Yurdugül, 2005). Çünkü kültürel faktörler, bireylerin teknolojiye yönelik tutumlarını ve kullanım biçimlerini etkileyebilmektedir (Olgun & Alatlı, 2021). Bu nedenle, Türk eğitim sisteminde yapay zekâ ile ilgili çalışmalara yön vermek ve yerel ihtiyaçlara uygun geçerli ve güvenilir ölçeklerin oluşturulmasını teşvik etmek amacıyla bu araştırma büyük önem taşımaktadır.

2. YÖNTEM

2.1. Araştırmanın modeli

Araştırmada eğitim öğretim süreçlerinde kullanılmak üzere yapay zekâ alanında geliştirilen ölçeklerin geçerlik ve güvenirliğini belirlemek için yapılan analizlerin yeterliliğini belirlemek amaçlandığından nitel araştırma desenlerinden meta-sentez yöntemi kullanılmıştır. Meta-sentez, belirli bir konu üzerinde yapılan araştırmaların tema veya ana şablonlar kullanılarak eleştirel bir bakış açısıyla sentezlenmesi ve yorumlanmasıdır (Çalık & Sözbilir, 2014). Bu doğrultuda eğitimde yapay zekâ kullanımına yönelik farklı araştırmacıların geliştirdikleri veya uyarladıkları ölçeklerin geliştirilme aşamaları derinlemesine incelenmiştir. Meta-sentez araştırmalarının işlem basamaklarını Polat ve Ay (2016) şu şekilde belirtmişlerdir:



Şekil 1. Meta-sentez araştırmalarının işlem basamakları

2.2. Araştırmanın çalışma grubu

Bu araştırma eğitimde yapay zekâ kullanımına yönelik geliştirilen ölçekler ile ilgili makale çalışmalarıyla sınırlıdır. Tarama yapılırken yayın yılı kısıtlaması yapılmamış ancak ulaşılan ölçek makalelerinin 2019-2023 yılları arasında yayınlanmış olduğu belirlenmiştir.

Araştırma sürecinde TR Dizin, EKUAL ve TOAD veri tabanlarından "yapay zekâ ölçeği", "yapay zekâ okuryazarlığı ölçeği", "yapay zekâ kaygı ölçeği" "yapay zekâ tutum ölçeği" ve "yapay zekâ öz yeterliği ölçeği" anahtar kelimeleri ile tarama yapılmış ve TR Dizin'de 6 makaleye, EKUAL'da 18 makaleye ve TOAD'da 12 ölçek makalesine ulaşılmıştır. İngilizce yayınlanan makaleler için "artificial intelligence scale", "artificial intelligence literacy scale", "artificial intelligence attitude scale" ve "artificial intelligence self-efficacy" anahtar kelimeleri ile yapılan taramada ERIC veri tabanında 77, EBSCOhost'da 391 makaleye ulaşılmıştır. Scopus'da 12984 ve Web of science'da 19544 makaleye erişim sağlanmış ve başlıklarında eğitim, yapay zekâ ve ölçek kelimeleri geçenler seçilmek suretiyle, başlık ile sınırlandırılarak Scopus'da 134 ve Web of science'da 136 makaleye indirgenmiştir. Ulaşılan makalelerin başlık ve özet kısımları incelenerek ölçek geliştirme analizleri içermeyen makaleler elenmiştir. Ayrıca birden fazla veri tabanında yer alan makaleler bir kez değerlendirmeye alınmıştır. Geliştirilen ölçekler için geçerlik ve güvenilirlik analizleri yapılmış olan makaleler belirlenmiştir. Tüm incelemeler sonucunda 16 tanesi Web of Science aynı zamanda Scopus, 3 tanesi Eric, 3 Ekual, 1 TrDizin ve 1 ise Ebsco veri tabanlarından olmak üzere toplam 24 adet makaleye ulaşılmıştır.

Yapılan tüm elemeler sonucunda; yapay zekâ okuryazarlığı ile ilgili 7, yapay zekâ kaygısı ile ilgili 4, yapay zekâ öz yeterliliği ile ilgili 2, yapay zekaya yönelik tutumla ilgili 2, yapay zekâ bağımlılığı, etiği, tehditleri, hazırlığı, kabulü, öz değerlendirmesi, kullanıma ilişkin görüşleri ve farkındalık düzeyleri ile ilgili birer ölçek makalesi ve ayrıca bir de meta-eğitim inanç ölçeği analizlere dahil edilmiştir.

Araştırmaya dahil edilen çalışmalar Ö1, Ö2, …, Ö24 şeklinde kodlanmış ve Tablo 1'de verilmiştir:

Tablo 1.

Araştırmaya Dâhil Edilen Çalışmalar

Kod	Yazar ve Yıl
Ö1	Dülger ve Köklü, (2023)
Ö2	Akkaya, Özkan ve Özkan, (2021)
Ö3	Kaya, Aydin, Schepman, Rodway, Yetişensoy ve Demir Kaya, (2024)
Ö4	Karaoğlan Yılmaz ve Yılmaz, (2023)
Ö5	Çelebi, Yılmaz, Demir ve Karakuş, (2023)
Ö6	Polatgil ve Güler, (2023)
Ö7	Karaca, Çalışkan ve Demir, (2021)
Ö8	Yilmaz, Yilmaz ve Ceylan, (2023)
Ö9	Ferikoğlu ve Akgün, (2022)
Ö10	Terzi, (2020)
Ö11	Suh ve Ahn, (2022)
Ö12	Laupichler, Aster, Perschewski ve Schleiss, (2023)
Ö13	Erol, Yurdakal ve Tekin Karagöz, (2023)
Ö14	Wang ve Chuang, (2024)
Ö15	Morales-García, Sairitupa-Sanchez, Morales-García ve Morales-García, (2024)
Ö16	Wang ve Wang, (2019)
Ö17	Jang, Choi ve Kim, (2022)
Ö18	Ng, Wu, Leung, Chiu ve Chu, (2024)
Ö19	Schepman ve Rodway, (2023)
Ö20	Chan ve Zhou, (2023)
Ö21	Wang, Rau ve Yuan, (2023)
Ö22	Hwang, Zhu ve Cui, (2023)
Ö23	Laupichler, Aster, Haverkamp ve Raupach, (2023)
Ö24	Morales-García, Sairitupa-Sanchez, Morales-García ve Morales-García, (2024)

2.3. Veri toplama araçları ve süreci

Ölçek geliştirme makalelerinin incelenmesi için araştırmacılar tarafından geliştirilen bir excel formu kullanılmıştır. Bu formda: incelenen makalenin adı, yazarları, yayın yeri ve tarihi, cilt-sayı numarası, ölçeğin geliştirilme amacı, benzer ölçeklerden farkı, örneklem büyüklüğü ve uygunluğu yanında geçerlik analizlerinin değerlendirilmesi için, ölçüt geçerliği, yüzey (görünüş) geçerliği, yapı (kavram) geçerliği, faktörleştirme tekniği, faktör isimlendirmesi, ayırt edicilik, kapsam (içerik) geçerliği (amaca hizmet etme düzeyi) bulunmakta; güvenirlik analizi olarak da kararlılık düzeyi ve iç tutarlılık düzeyi bilgileri yer almaktadır.

2.4. Verilerin analizi

Araştırma kapsamında değerlendirilen ölçek geliştirme çalışmaları betimsel analize tabi tutularak kategorik sınıflandırmalar yapılmıştır. Betimsel analiz, tümdengelimci bir analiz türü olup, araştırmanın kavramsal yapısının önceden belirlendiği araştırmalarda kullanılır (Yıldırım & Şimşek, 2021). Çalışma kapsamında toplanan veriler betimsel olarak analiz edilerek frekans değerleri tablolaştırılmıştır. İlk olarak kategorilerin sınıflandırılmasını kolaylaştırmak amacıyla bir tablo oluşturulmuş ve bir ölçme ve değerlendirme uzmanının görüşüne sunulmuş. İkinci aşamada iki araştırmacı tarafından bu form kullanılarak araştırtma kapsamına alına makalelerden beş tanesi birlikte değerlendirilerek kriterlere ilişkin fikir birliği sağlanmıştır. Daha sonra tüm makaleler iki araştırmacı tarafından bağımsız olarak değerlendirilmiştir. Son aşamada her iki değerlendirme karşılaştırılarak görüş farklılıkları olan kısımlar yeniden incelenmiş ve görüş birliği sağlanarak verilerin geçerlilik ve güvenilirliği sağlanmıştır.

3. BULGULAR

Araştırma bulgularına göre dört tema oluşturulmuştur. Bunlar araştırma amacı, örneklem, geçerlik ve güvenirliktir. Belirlenen temalar doğrultusunda tablolar oluşturulmuş ve yorumlanmıştır. Eğitimde yapay zekâ kullanımına ilişkin geliştirilmiş olan ölçeklerin amaçları Tablo 2'de verilmiştir.

Tablo 2.
Ölçeklerin Türü ve Geliştirilme Amaçları

Ölçek Türü	Ölçek Amacı	Ölçek Kodu	f	
Geliştirme	Okuryazarlık	Ö18, Ö21, Ö22, Ö23	4	16
	Tutum	Ö11, Ö17, Ö19	3	
	Algı	Ö7, Ö20	2	
	Kabul, İnanç	Ö8, Ö13	2	
	Kaygı	Ö16	1	
	Özyeterlik	Ö14	1	
	Farkındalık	Ö9	1	
	Bağımlılık	Ö24	1	
	Kullanım	Ö1	1	
Uyarlama	Okuryazarlık	Ö4, Ö5, Ö6, Ö12	4	8
	Tutum	Ö3	1	
	Algı	Ö2	1	
	Kaygı	Ö10	1	
	Özyeterlik	Ö15	1	

Tablo 2 incelendiğinde eğitimde yapay zekâ kullanımına ilişkin geliştirilen ölçeklerin Türk kültürüne uyarlama (n = 8) ve geliştirme (n = 16) şeklinde tasarlandıkları görülmüştür. Ölçeklerin geliştirilme amaçları incelendiğinde okuryazarlık (n = 8), tutum (n = 4), kaygı (n = 2), algı (n = 3), özyeterlik (n = 2) ve birer adet kabul, inanç, farkındalık, kullanım ve bağımlılık konulu ölçeklerin geliştirildiği belirlenmiştir. Buna göre alanyazında en çok eğitimde yapay zekâya dönük okuryazarlık ve tutum değişkenleri üzerinde durulduğunu söylemek mümkündür. Eğitimde yapay zekâ kullanımına ilişkin geliştirilmiş olan ölçeklerin çalışma grubu, örneklem büyüklüğü ve uygunluğu Tablo 3'te verilmiştir.

Tablo 3.
Ölçeklerin Örneklem Türü ve Örneklem Büyüklüğü

Örneklem	Ölçek kodu	f	
K12 öğrencileri	Ö11, Ö18,	2	24
Üniversite öğrencileri	Ö2, Ö7, Ö8, Ö12, Ö13, Ö15, Ö17, Ö20, Ö22, Ö24,	10	
Öğretmen	Ö1, Ö9, Ö10,	3	
Yetişkin	Ö3, Ö4, Ö5, Ö6, Ö14, Ö16, Ö19, Ö21, Ö23,	9	
Örneklem büyüklüğü	**Ölçek kodu**	**f**	
0-100 arasında	Ö12,	1	24
201-300 arasında	Ö1, Ö13,	2	
301-400 arasında	Ö3, Ö10, Ö11, Ö14, Ö16, Ö18, Ö19, Ö22,	8	
401-500 arasında	Ö2, Ö5, Ö15, Ö20, Ö23,	5	
501-600 arasında	Ö6, Ö9, Ö24,	3	
601-700 arasında	Ö4, Ö8, Ö21,	3	
801-900 arasında	Ö7,	1	
1001 ve üzeri	Ö17,	1	

Tablo 3'te görüldüğü üzere ölçek geliştirme çalışmalarında K12 (ilkokul, ortaokul ve lise) ve üniversite öğrencileri, öğretmenler ve lise ve üzeri yaş grubu yetişkinler tercih edilmiştir. Çalışma grubu olarak en çok üniversite öğrencileri (n = 10), en az ise K12 öğrencileri (n = 2) ile çalışılmıştır. Örneklem büyüklüğüne göre ölçek geliştirme çalışmaları incelendiğinde çok farklı örneklem büyüklüklerinin seçildiği görülmektedir. En çok tercih edilen örneklem büyüklüğü 301-400 arası (n = 8) iken 100'den az, 800 ve 1000'den fazla birer çalışma olmuştur. Buna göre K4 ve K8 düzeyinde hiç ölçek olmadığını, geliştirilen ölçekleri büyük çoğunluğunun üniversite öğrencilerine, öğretmenlere ve yetişkinlere dönük olduğu söylenebilir. Örneklem büyüklükleri incelendiğinde ise geliştirilen ölçeklerin büyük bölümünün 300'ün üzerinde olduğunu söylemek mümkündür. Eğitimde yapay zekâ kullanımına ilişkin geliştirilmiş olan ölçeklerin geçerlik analizleri Tablo 4'te verilmiştir.

Tablo 4.
Ölçeklerin Geçerlik Analizleri

Geçerlik analizi		**Ölçek kodu**	**f**	
Yüzey (görünüş) geçerliği	3-5 uzman	$Ö_3$, $Ö_5$, $Ö_9$, $Ö_{10}$, $Ö_{13}$, $Ö_{14}$, $Ö_{21}$,	7	19
	6-8 uzman	$Ö_1$, $Ö_2$, $Ö_6$, $Ö_7$, $Ö_8$, $Ö_{11}$, $Ö_{12}$, $Ö_{16}$, $Ö_{17}$, $Ö_{18}$,	10	
	9 ve üzeri	$Ö_{15}$, $Ö_{23}$,	2	
Kapsam (içerik) geçerliği	Kapsam geçerlik oranı (KGO=CVR=Content Validity Ratio)	$Ö_{11}$,	1	2
	Kapsam geçerlik indeksi (KGİ=CVI= Content Validity Index)	$Ö_{17}$,	1	
Kriter (Ölçüt) Geçerliği	Benzer çalışma tutarlılığı	$Ö_{16}$, $Ö_{17}$, $Ö_{19}$,	3	3
Yapı (kavram) geçerliği	AFA	$Ö_{13}$, $Ö_{16}$, $Ö_{23}$,	3	22
	DFA	$Ö_1$, $Ö_2$, $Ö_3$, $Ö_4$, $Ö_5$, $Ö_{15}$, $Ö_{18}$, $Ö_{19}$, $Ö_{20}$, $Ö_{22}$, $Ö_{24}$,	11	
	AFA ve DFA	$Ö_6$, $Ö_7$, $Ö_8$, $Ö_9$, $Ö_{10}$, $Ö_{11}$, $Ö_{17}$, $Ö_{21}$,	8	
Madde ayırt ediciliği	Madde-toplam korelasyonu	$Ö_8$, $Ö_9$, $Ö_{13}$, $Ö_{14}$, $Ö_{15}$, $Ö_{16}$, $Ö_{17}$, $Ö_{21}$, $Ö_{24}$,	9	11
	Üst ve alt %27 karşılaştırması	$Ö_4$, $Ö_8$,	2	

Tablo 4'te görüldüğü üzere geçerlik analizleri için yüzey, kapsam, kriter ve yapı geçerlik analizleri kullanılmıştır. Bu analizlerden en fazla yapı geçerliği (n = 22), en az ise ölçüt geçerliği (n = 3) kullanılan analizler arasındadır. Yüzey geçerliği için çoğunlukla (n=19) uzman görüşleri alınmış ancak sadece iki çalışmada uzman görüşleri arasındaki uyum istatistiklerinin bilgisi verilmiştir. İncelenen makalelerin %92'sinde yapı geçerliği için AFA veya DFA analizlerinden en az birinin yapıldığı görülmektedir. Analizlerin yarısında (n=11) DFA analizi tek başına tercih edilirken; üçte birinde (n=8) AFA ve DFA analizleri birlikte kullanılmıştır. Çalışmaların %79'unda DFA analizinin tercih edilmesinin nedeni olarak, incelenen makalelerin üçte birinin (n=8) uyarlama çalışması olması söylenebilir.

Madde ayırt ediciliği için dokuz çalışmada madde-toplam korelasyonu hesaplanmış ve sadece iki çalışmada alt-üst grup farklılaşmasının analiz edildiği belirlenmiştir. Araştırmacılar tarafından, ölçek geliştirme aşamalarında madde ayırt ediciliği geçerliğine %46 oranında dikkat edildiği söylenebilir. Geliştirilmiş olan ölçeklerin güvenirlik analizleri Tablo 5'te verilmiştir.

Tablo 5.
Ölçeklerin Güvenirlik Analizleri

Güvenirlik analizi			**Ölçek kodu**	**f**	
İç tutarlılık	Cronbach alfa	$0.90 \leq \alpha$	$Ö_1$, $Ö_2$, $Ö_4$, $Ö_6$, $Ö_8$, $Ö_9$, $Ö_{10}$, $Ö_{11}$, $Ö_{13}$, $Ö_{14}$, $Ö_{15}$, $Ö_{16}$, $Ö_{18}$, $Ö_{23}$, $Ö_{24}$,	15	24
		$0.80 \leq \alpha < 0.90$	$Ö_3$, $Ö_5$, $Ö_7$, $Ö_{12}$, $Ö_{19}$, $Ö_{21}$,	6	
		$0.70 \leq \alpha < 0.80$	$Ö_{17}$, $Ö_{20}$, $Ö_{22}$,	3	
	McDonald omega	$0.90 \leq \omega$	$Ö_{15}$,	1	2
		$0.80 \leq \omega < 0.90$	$Ö_{24}$	1	
Test tekrar test			$Ö_4$, $Ö_6$, $Ö_8$,	3	3

Tablo 5'te görüldüğü üzere güvenirlik analizleri için iç tutarlılık ve test tekrar test analizleri tercih edilmiştir. İncelenen çalışmaların tamamında iç tutarlılık analizi için Cronbach alfa katsayısı tercih edilirken iki çalışmada ek olarak McDonald omega katsayısının hesaplandığı belirlenmiştir. İncelenen makalelerin %63'ünde toplam ve faktörleri bazında güvenirlik katsayısı 0.90 ve üzerinde hesaplanırken, McDonald omega katsayısı 0.87 ve üzerinde hesaplanmıştır. Çalışmaların büyük çoğunluğunda Cronbach alfa ve McDonald omega katsayısı 0.70 ve üzerinde güvenirlik düzeyi belirtilirken sadece üç ölçeğin birer faktöründe Cronbach alfa katsayı 0.60 ile 0.70 aralığında hesaplanmıştır. Bununla birlikte sadece üç ölçek geliştirme çalışmasında kararlılık analizi yapılarak, test tekrar test analizi kullanılmıştır. Araştırmacıların ölçek geliştirme çalışmalarında genel olarak kararlılık analizini önemsemedikleri söylenebilir.

4. TARTIŞMA ve SONUÇ

Eğitimde yapay zekâ kullanımına yönelik geliştiren ölçeklerin geçerlik ve güvenirlik çalışmalarının uygunluğunun değerlendirildiği bu çalışmada ilk olarak ölçeklerin türü ve geliştirilme amaçları incelenmiştir. Bu alanda geliştirilen ölçeklerin üçte birinin farklı dillerden Türk kültürüne uyarlama şeklinde hazırlandığı; üçte ikisinin ise bağımsız olarak geliştirildiği görülmektedir. İncelenen 24 ölçeğin geliştirilme amacının çoğunlukla okuryazarlık ve tutum konularında olduğu görülmektedir. Okuryazarlık konusu incelediğimiz ölçeklerde çokça tercih edilen bir konu olmasına rağmen alanyazında, çalışmamızın aksine az tercih edilen bir konu olduğu ifade edilerek, bu konudaki araştırmaların arttırılması gerektiği belirtilmiştir (Cin Şeker & Yücel Çetin, 2022; Konu Kadirhanoğulları, 2024). Bunun yanında araştırmada ortaya çıkan ölçek geliştirme ve uyarlama çalışmalarında tutum ölçeği geliştirmenin çoğunlukla tercih edilmiş olduğu sonucu Ergene (2020) ve Konu Kadirhanoğulları (2024) çalışmaları ile örtüşmektedir. Kullanım alanlarının genişliğinden dolayı tutum ölçeklerinin kullanımı daha fazla tercih edilmektedir (Cin Şeker & Yücel Çetin, 2022). Tutum ölçeklerini kaygı, algı ve özyeterlik konuları takip etmektedir. Cin Şeker ve Yücel Çetin (2022) çalışmasında tutum ölçeklerini öz yeterlik, motivasyon ve kaygı ölçme amacıyla geliştirilen ölçeklerin izlemesi çalışmamızla benzerlik göstermektedir. Kabul, inanç, kullanım, farkındalık ve bağımlılık konulu ölçek geliştirme çalışmaları daha az tercih edilmiştir. Alanyazı ile karşılaştırıldığında bağımlılık (Olgun & Alatlı, 2021), inanç (Cin Şeker & Yücel Çetin, 2022; Ergene, 2020), yetki kullanım kaygısı (Ayyıldız, 2022) ve bilişsel farkındalık (Cin Şeker & Yücel Çetin, 2022) konularında az sayıda ölçek geliştirilmesi araştırmamızla benzerlik göstermektedir. Gelecekte yapay zekâ teknolojilerinin öğretmenlerin yerini alma kaygısı, bu teknolojilerin eğitime entegre edilmesi önündeki engeller arasında bulunmaktadır (Eyüp & Kayhan, 2023). Bu durumda yapay zekâ kullanım kaygısı konusunda ölçek geliştirme çalışmalarının artması gerekliliği dikkat çekmektedir.

Eğitimde yapay zekâ kullanımına ilişkin geliştirilen ölçeklerde ikinci inceleme konusu çalışma grubu, örneklem büyüklüğü ve uygunluğudur. Eğitimde yapay zekâ kullanımına yönelik ölçek geliştirme araştırmaları için çalışma grubu olarak en çok üniversite öğrencileri tercih edilmiştir. Lise ve üzeri yaş grubundan oluşan yetişkinler de yine çoğunlukla tercih edilen çalışma grubunu oluşturmaktadır. Benzer olarak Fen ve Matematik Eğitimi alanında gerçekleştirilen ölçek geliştirme araştırmalarında (Gül & Sözbilir, 2015) en fazla lisans öğrencileri tercih edilmiş; çalışmamızın aksine üniversite öğrencileri ile eğitim yönetimi alanında (Ayyıldız, 2022) ve okuma eğitimi alanında (Cin Şeker & Yücel Çetin, 2022) az sayıda ölçek geliştirme çalışmaları gerçekleştirilmiştir. Eğitimde yapay zekâ kullanımına yönelik öğretmenlere özel geliştirilmiş ölçekler de literatürde mevcut olup, biyoloji eğitiminde (Konu Kadirhanoğulları, 2024), Fen ve Matematik Eğitimi alanında (Gül & Sözbilir, 2015) ve eğitim yönetimi alanında (Ayyıldız, 2022) eğitimcilere yönelik az sayıda ölçek geliştirme çalışmaları olduğu ifade edilmektedir. Eğitimde yapay zekâ kullanımına yönelik ölçek geliştirmede çalışma grubu olarak en az K12 (ilkokul, ortaokul ve lise) öğrencileri tercih edilmiştir. Aksine alanyazında okuma eğitimi alanında (Cin Şeker & Yücel Çetin, 2022) ve matematik eğitimi alanında (Ergene, 2020) ölçek geliştirmek için en fazla ortaokul öğrencileri ile biyoloji eğitimi alanında (Konu Kadirhanoğulları, 2024) ise K12 ile üniversite öğrencileri ile çalışılmıştır. Eğitim öğretimin en önemli paydaşlarından olan ilkokul ve ortaokul öğrencilerine yönelik yapay zekâ kullanımı ile ilgili neredeyse hiç ölçek geliştirilmemiş olması önemli bir eksiklik olarak dile getirilebilir. Bu durumda araştırmacıların eğitimin temel elemanlarından olan K4, K8 ve K12 öğrencilerine yönelik eğitimde yapay zekâ kullanımına yönelik ölçek geliştirme çalışmalarını arttırması, bir ihtiyaç olarak görünmektedir.

Geliştirilen ölçeklerin örneklem büyüklüğü incelendiğinde çalışmaların %80'den fazlasının 300'den büyük olduğu ve yeterli olduğu görülmektedir. Eminoğlu Özmercan ve Kemer (2024) ölçek geliştirme çalışmaları incelenen tezlerin yarısının tamamen uygun, üçte birinin ise kısmen uygun olduğunu; Ayyıldız (2022) üçte ikisinin 300 ve üzeri katılımcı sayısına sahip olduğu ifade etmiştir. Olgun ve Alatlı (2021) ölçek geliştirme çalışmalarının üçte birinde madde başına düşen katılımcı sayısının 20 veya daha fazla olduğunu ifade etmektedir. Kahraman, Aytekin ve Alparslan (2021) ölçek geliştirme çalışmalarının yine üçte birinin (%30.8) uygun örneklem büyüklüğü ile çalışıldığını; %15'inin ise örneklem büyüklüğünün belirtilmediğini ifade etmiştir. Ergene (2020) incelenen çalışmaların dörtte üçünden fazlasının uygun örneklem büyüklüğü ile çalışıldığını; Gül ve Sözbilir (2015) incelen makalelerin yaklaşık olarak yarısının uygun büyüklükte olduğunu, dörtte birinin ise uygun örneklem sayısının altında kaldığını ifade etmiştir.

Eğitimde yapay zekâ kullanımına yönelik geliştirilen ölçeklerin dörtte birinde katılımcı sayısı madde sayısının 10 katından az; sekizde birinde 300'den az katılımcı ile çalışılmıştır. Ölçek geliştirme çalışmalarının dörtte üçü gibi büyük bir oranında örneklem büyüklüğünün madde sayısına oranla yeterli olduğu söylenebilir. Örneklem büyüklüğü, ölçek geliştirme sürecinde önemli bir faktördür, çünkü güvenilir ve geçerli sonuçlar elde etmek için yeterli sayıda katılımcıya ihtiyaç vardır (DeVellis, 2014). DeVellis (2014), örneklem büyüklüğünün, kullanılan istatistiksel yöntemlerin gücünü etkilediğini ve bu nedenle örneklem sayısının belirlenmesinde dikkatli olunması gerektiğini vurgulamaktadır. Ayrıca, karmaşık istatistiksel analizler için örneklem büyüklüğünün genellikle 200-300 arasında olması gerektiğini önermektedir. Erkuş (2012), özellikle psikolojik testlerde, geçerli sonuçlar elde edebilmek için örneklemin homojen olmaması durumunda daha büyük örneklem boyutlarına ihtiyaç duyulacağını belirtmektedir ve genellikle 100-200 katılımcı önerilmektedir. Kılıç (2023) ise, R programlama dili kullanılarak yapılan ölçek geliştirme çalışmalarında örneklem büyüklüğünün, modelin doğruluğunu ve parametre kestirimlerinin güvenilirliğini etkilediğini ifade ederek, karmaşık modeller için örneklem sayısının 300'ün üzerine çıkmasının tercih edilebileceğini vurgulamaktadır. Benzer şekilde Olgun ve Alatlı (2021) ölçek geliştirme çalışmalarında %12,5 oranında uygun olmayan örneklem büyüklüğüyle çalışıldığını, Eminoğlu Özmercan ve Kemer (2024) ölçek geliştirme çalışmaları incelenen tezlerin %16'sının hiç uygun olmadığını ifade etmiştir.

Eğitimde yapay zekâ kullanımına ilişkin geliştirilen ölçeklerin üçüncü teması geçerlik analizleridir. En fazla tercih edilen geçerlik türü yapı geçerliği olmuştur. Yapı geçerliliği, bir ölçüm aracının amacına uygun olarak ölçmek istediği kavramı doğru şekilde yansıtıp yansıtmadığını belirlemede kritik bir rol

oynamaktadır (DeVellis, 2014). Bu geçerliliği sağlamak için faktör analizleri gibi istatistiksel yöntemler büyük önem taşır ve testin yapısının doğru şekilde tanımlanması gerektiğini vurgular (Erkuş, 2012). Bilgisayar teknolojilerinin analizlerde kullanılmasıyla birlikte 2000'li yıllardan itibaren AFA ve DFA tercihlerinde artış görülmüş ve bu durum yapı geçerliğinin daha çok tercih edilmesine neden olmuştur (Erkuş, 2019; Yılmaz & Özgül, 2023). İncelenen ölçek geliştirme çalışmaların yarısına yakınında doğrulayıcı faktör analizi (DFA) tercih edilmiş; üçte birinde ise doğrulayıcı ve açımlayıcı faktör analizleri (AFA) birlikte kullanılmıştır. AFA ve DFA'nın birlikte test edildiği çalışmaların dörtte birinde analizler için farklı örneklem kullanıldığı belirtilmemiştir. Literatürde yapı geçerliği için çoğunlukla AFA'nın (Acar Güvendir & Özer Özkan, 2015; Gül & Sözbilir, 2015; Konu Kadirhanoğulları, 2024; Savaş & Pazar, 2021) veya AFA ile birlikte DFA'nın kullanıldığı (Ergene, 2020; Kahraman vd., 2021; Şahin & Boztunç Öztürk, 2018) ancak çalışmaların yarısına yakınında bu analizlerin aynı veriler kullanılarak yapıldığı ifade edilmiştir (Olgun & Alatlı, 2021). AFA var olan psikolojik yapıyı anlamlandırırken, DFA bu yapıyı test etmektedir (Erkuş, 2019) öyleyse temel yapıyı belirlemek için analizlere AFA ile başlanarak farklı bir örneklem kullanılmasıyla DFA ile devam edilmelidir (Cabrera-Nguyen, 2010; Gül & Sözbilir, 2015). Oysaki incelenen ölçek geliştirme çalışmalarının sadece yarısına yakınında AFA veya AFA ile DFA'nın birlikte kullanıldığı görülmektedir. Araştırmamızda DFA analizlerinin çoğunlukta olması incelenen makalelerin üçte birinin uyarlama çalışması olmasından kaynaklanabilir. Bununla birlikte çalışmaların %46'sında madde ayırt ediciliği analizleri yapılmıştır. Araştırmacılar ayırt edicilik geçerliğini genel itibariyle göz ardı etmektedirler (Gül & Sözbilir, 2015). Oysaki yapıyı oluşturan değişkenler ile diğer değişkenler arasındaki düşük korelasyon gösterilerek yapının geçerliği ile ayırt edicilik geçerliğinin sağlanması gerekmektedir (Churchill, 1979).

Yapı geçerliğinden sonra en fazla yüzey geçerliği analizleri tercih edilmiştir. Yüzey geçerliği, bir ölçüm aracının, ölçtüğü kavramla ne kadar uyumlu olduğunu değerlendiren bir geçerlik türüdür. DeVellis (2014), yüzey geçerliğinin, testin kullanıcıları tarafından anlaşılabilir ve anlamlı bulunması gerektiğini belirtmektedir. Bu, testin tasarımında, içeriğin amacına uygun ve anlaşılır olması açısından önemlidir. Erkuş (2012), yüzey geçerliğinin, testin uygulandığı grup için ilk izlenimlerin doğruluğunu yansıttığını ancak bu tür geçerliğin tek başına yeterli olmadığını ifade etmektedir. İncelenen çalışmaların yaklaşık olarak dörtte birinde alınan uzman görüşü ile ilgili bilgi verilmezken; yaklaşık olarak yarısında 6-8 alan uzmanının görüşlerine başvurulmuştur. Çoğunlukla ilgili alan ve dil uzmanı görüşleri alınmış olup sadece bir çalışmada ölçme değerlendirme uzmanından görüş alınmıştır. Çalışmaların yalnızca birinde (Ö11) kapsam geçerlik oranı şeklinde istatistiki bilgi verilmiş, diğer tüm çalışmalarda yalnızca uzman görüşlerine başvurulduğu ifade edilmiştir. Benzer şekilde yüzey geçerliği için istatistik yerine sadece uzman görüşü alınması Gül ve Sözbilir (2015) çalışmasında ifade edilmiştir. Alanyazında yine yapılan ölçek geliştirme çalışmalarında en az yüzey geçerliğinin incelendiği ifade edilmektedir (Ayyıldız, 2022; Konu Kadirhanoğulları, 2024; Yılmaz & Özgül, 2023).

İncelenen ölçek geliştirme çalışmaların sadece beşte ikisinde amaca hizmet etme düzeyini belirlemek amacıyla madde toplam korelasyonu analizi yapılmıştır. Ölçek maddelerinin ölçülmek istenen yapıyı kapsayacak nitelikte olması kapsam geçerliğidir (Ergene, 2020). Literatürde kapsam geçerliği ile görünüş geçerliği kavramlarının karıştırıldığı görülmektedir (Ergene, 2020; Olgun & Alatlı, 2021; Şahin & Boztunç Öztürk, 2018). Ölçek geliştirildikten sonra sınanması gereken görünüş geçerliğinin genellikle ilk aşamalarda yapılmasının bu kavram karmaşasına neden olabileceği ifade edilmiştir (Gül & Sözbilir, 2015). Oysaki kapsam geçerliği, uzman görüşüne dayalı nitel yapının istatistiksel sonuçlara dönüştürülmesi olarak ifade edilmektedir (Yurdugül, 2005).

Ölçüt geçerliği, bir testin dış bir ölçümle ne kadar ilişkilendiğini ve bu ilişkinin gücünün testin doğruluğunu belirlemede kritik bir rol oynadığını gösterir (DeVellis, 2014). Bu doğrultuda, bir testin doğruluğunu değerlendirmek için benzer bir test veya ölçütle karşılaştırılması önemlidir (Erkuş, 2012). İncelenen çalışmalarda en az (%12,5) ölçüt geçerliği yapıldığı görülmüştür. Benzer durum alanyazında

görülmekte ve ölçüt geçerliğine yeterince önem verilmediği ifade edilmektedir (Ergene, 2020; Gül & Sözbilir, 2015; Kahraman vd., 2021; Konu Kadirhanoğulları, 2024; Şahin & Boztunç Öztürk, 2018; Savaş & Pazar, 2021; Yılmaz & Özgül, 2023). Erkuş (2017)'a göre yapı geçerliği sadece DFA ve AFA ile incelenemez, güvenilir ve geçerliği kanıtlanmış benzer ölçeklerler korelasyonu mutlaka incelenmelidir.

Son tema olarak güvenirlik analizleri incelendiğinde geliştirilen veya uyarlanan ölçeklerin %60'ında güvenirlik oranı .90'dan fazla ölçülüp, kalanında .70 üzerinde hesaplandığı ve çalışmaların Cronbach alfa katsayısının hesaplandığı ancak sadece iki çalışmada McDonald omega katsayısının hesaplandığı görülmüştür. Cronbach alfa (α), iç tutarlılığı ölçen en yaygın kullanılan güvenirlik katsayısıdır. DeVellis (2014) ve Erkuş (2012), Cronbach alfa değerinin genellikle .70 ve üzeri olması gerektiğini belirtmektedir. .70'in altındaki değerler, testin güvenilirliğini sorgulanabilir kılabilir. Bunun yanında McDonald omega katsayısı, iç tutarlılığı ölçmek için kullanılan diğer bir güvenirlik katsayısıdır ve özellikle karmaşık yapılar için daha güvenilir bir alternatif olarak öne çıkmaktadır. Erkuş (2012), bu katsayının, Cronbach alfa'ya göre daha esnek olduğunu ve testin faktör yapısına daha iyi uyum sağladığını ifade etmektedir. İncelenen çalışmaların tamamında Cronbach alfa katsayısının hesaplanma durumu, Gül ve Sözbilir (2015), Ayyıldız (2022), Acar Güvendir ve Özer Özkan (2015) çalışmalarında görülürken, Gül ve Sözbilir (2015) incelenen makalelerin tamamında Cronbach alfa katsayısının kullanılmasını doğru bir tercih olarak belirtmiştir. Tamamına yakın (%90'dan fazla) oranda Cronbach alfa katsayısı hesaplanan çalışmalar da çoğunluktadır (Kahraman vd., 2021; Olgun & Alatlı, 2021). Cronbach alfanın hesaplanması için hazır programların olması daha çok tercih edilmesine neden olmaktadır (Şahin & Boztunç Öztürk, 2018). Güvenirlik analizi için iç tutarlılık hesaplamalarında Cronbach alfa katsayısının tercih edilmesi literatür ile uyum göstermektedir.

Test tekrar test güvenirliği, aynı testi farklı zamanlarda uygulayarak elde edilen sonuçların tutarlılığını ölçer. Bu analizde genellikle .80 veya daha yüksek bir korelasyon değeri istenir (Erkuş, 2012). İncelenen ölçek geliştirme makalelerinin sadece sekizde birinde test tekrar test analizi ile ölçeğin kararlılığının hesaplandığı görülmüştür. Literatürde test tekrar test yönteminin kullanımının %15'in altında kaldığı birçok çalışma bulunmaktadır (Ayyıldız, 2022; Gül & Sözbilir, 2015; Konu Kadirhanoğulları, 2024; Savaş & Pazar, 2021). Test tekrar test yönteminin uygulanması aynı formun iki defa uygulanmasını gerektirdiğinden, araştırmacılar tarafından emek ve zaman israfı olarak görülmekte ve tek uygulamaya dayalı Cronbach alfa katsayısı hesaplamasına yönelim artmaktadır (Gül & Sözbilir, 2015). Bu çalışma kapsamında elde edilen sonuçlar bu açıdan literatür ile uyumlu görünmektedir. Ancak literatürde Cronbach alfa katsayısı ile test tekrar test yönteminin birlikte kullanılması yaygıdır (Ergene, 2020; Kahraman vd., 2021; Olgun & Alatlı, 2021; Yılmaz & Özgül, 2023).

5. ÖNERİLER

Bu araştırmada, eğitimde yapay zekâ kullanımına yönelik geliştirilen ölçeklerin türü, amacı, çalışma grubu, örneklem büyüklüğü, geçerlik ve güvenirlik analizleri betimsel olarak incelenmiştir. Bulgular, ölçeklerin büyük oranda tutum ve okuryazarlık konularına odaklandığını, ancak okuryazarlık konusundaki araştırmaların alanyazında sınırlı kaldığını göstermektedir. Çalışma grubu olarak en fazla üniversite öğrencilerinin tercih edildiği, ancak K12 seviyesindeki öğrenciler için ölçek geliştirme çalışmalarının yetersiz olduğu tespit edilmiştir. Örneklem büyüklüğünün genellikle yeterli olduğu, ancak bazı çalışmaların madde başına düşen katılımcı sayısı açısından yetersiz kaldığı görülmüştür. Geçerlik analizlerinde yapı geçerliğinin en yaygın kullanılan yöntem olduğu, ancak AFA ve DFA'nın birlikte test edildiği çalışmaların yalnızca yarısında farklı örneklemler kullanıldığı belirlenmiştir. Ayrıca, kapsam ve ölçüt geçerliği gibi kritik geçerlik türlerine yeterince önem verilmediği görülmektedir. Güvenirlik analizlerinde Cronbach alfa katsayısının yaygın olarak kullanıldığı, ancak test tekrar test yönteminin düşük oranda tercih edildiği saptanmıştır. Bu doğrultuda, gelecekteki araştırmalarda okuryazarlık, bağımlılık ve yetki kullanım kaygısı gibi konulara yönelik ölçek geliştirme çalışmalarının artırılması, K12 düzeyinde öğrencilere yönelik ölçekler oluşturulması ve ölçek geçerliği için farklı yöntemlerin daha yaygın kullanılması önerilmektedir. Ayrıca, test tekrar test analizinin daha sık uygulanması, ölçeklerin zaman içindeki kararlılığını değerlendirme açısından faydalı olacaktır.

Kaynakça/Reference

Acar Güvendir, M., & Özer Özkan, Y (2015). Türkiye'deki eğitim alanında yayımlanan bilimsel dergilerde ölçek geliştirme ve uyarlama konulu makalelerin incelenmesi. *Elektronik Sosyal Bilimler Dergisi*, *14*(52), 23-33. doi: 10.17755/esosder.54872

Acar Güvendir, M. ve Özer Özkan, Y. (Editörler). (2022). *Tüm yönleriyle ölçek geliştirme süreci.* Pegem Akademi.

Akkaya, B., Özkan, A. & Özkan, H. (2021). Yapay zekâ kaygı (YZK) ölçeği: Türkçeye uyarlama, geçerlik ve güvenirlik çalışması. *Alanya Akademik Bakış,* 5(2), 1125-1146. https://doi.org/10.29023/alanyaakademik.833668

Ayyıldız, P. (2022). Eğitim yönetimi alanında "liderlik, okul yönetimi, okul müdürü" temalı ölçek geliştirme/uyarlama çalışmalarının tematik içerik analizi ile incelenmesi, *International Journal of Eurasia Social Sciences (IJOESS)*, 13(49), 968-983. Doi: 10.35826/ijoess.3189

Başol, G. (2019). *Eğitimde ölçme ve değerlendirme. (6. Baskı).* Pegem Akademi.

Büyüköztürk, Ş. (2023). *Sosyal bilimler için veri analizi el kitabı*. Pegem Akademi.

Cabrera-Nguyen, P. (2010). Author guidelines for reporting scale development and validation results in the Journal of the Society for Social Work and Research. *Journal of the Society for Social Work and Research*, *1*(2), 99-103. Doi: 10.5243/jsswr.2010.8

Chan, C. K. Y. & Zhou, W. (2023). An expectancy value theory (EVT) based instrument for measuring student perceptions of generative AI. *Smart Learn. Environ.* 10(1), 64. https://doi.org/10.1186/s40561-023-00284-4

Chan, K. S. & Zary, N. (2019). Applications and challenges of ımplementing artificial ıntelligence in medical education: Integrative review. *JMIR Medical Education*, *5*(1), e13930. doi: 10.2196/13930

Chen, L., Chen P. & Lin, Z. (2020). Artificial ıntelligence in education: A review, *in IEEE Access*, 8, 75264-75278. doi: 10.1109/ACCESS.2020.2988510

Chen, X., Xie, H., Zou, D. & Hwang, G. J. (2020). Application and theory gaps during the rise of artificial ıntelligence in education. *Computers and Education: Artificial Intelligence*, 1 (July),1-20, 100002. https://doi.org/10.1016/j.caeai.2020.100002

Chen, X., Zou, D., Xie, H., Cheng, G. & Liu, C. (2022). Two decades of artificial ıntelligence in education. *Educational Technology & Society*, *25*(1), 28-47. https://www.jstor.org/stable/48647028

Churchill, G. A. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research*, *16*(1), 64–73. https://doi.org/10.2307/3150876

Cin Şeker, Z. ve Yücel Çetin, D. (2022). Okuma eğitimi alanında yapılan ölçek geliştirme çalışmalarının incelenmesi: Bir meta-sentez çalışması. *Uluslararası Türkçe Edebiyat Kültür Eğitim Dergisi, 11*(3), 1193- 1209.

Çalık, M. & Sözbilir, M. (2014). İçerik analizinin parametreleri. *Eğitim ve Bilim*, *39* (174), 33-38. http://dx.doi.org/10.15390/EB.2014.3412

Çelebi, C., Yılmaz, F., Demir, U. & Karakuş, F. (2023). Artificial ıntelligence literacy: An adaptation study. *Instructional Technology and Lifelong Learning*, 4(2), 291-306. https://doi.org/10.52911/itall.1401740

Çelen, Ü., & Aybek, E. C. (2022). A novel approach for calculating the item discrimination for Likert type of scales. *International Journal of Assessment Tools in Education*, 9(3), 772-786. https://doi.org/10.21449/ijate.1173356

Çoban, E. & Uzun H. (2022). Endüstri 4.0'ın eğitim alanına etkileri. *Fırat Üniversitesi Uluslararası İktisadi ve İdari Bilimler Dergisi*, *6*(1), 97-124.

Cukurova, M., Luckin, R. & Kent, C. (2020). Impact of an artificial intelligence research frame on the perceived credibility of educational research evidence. *International Journal of Artificial Intelligence in Education*, *30*(2), 205-235. https://doi.org/10.1007/s40593-019-00188-w

Demirel, Ö. (2020). *Eğitimde program geliştirme: Kuramdan uygulamaya. (29. Baskı).* Pegem Akademi.

DeVellis, R. F. (2014). *Ölçek geliştirme: Kuram ve uygulamalar*. Nobel Akademik Yayıncılık.

DeVellis, R. F. (2017). *Scale development: Theory and applications (4th ed.).* Sage Publications.

Dülger, E. D. & Köklü, M. (2023). Okul yöneticilerinin ve öğretmenlerin eğitimde yapay zekâ kullanımına ilişkin görüşlerini belirlemeye yönelik bir ölçek geliştirme çalışması. *ISPEC International Journal of Social Sciences & Humanities*, 7(1), 154-174. https://doi.org/10.5281/zenodo.7767140

Eminoğlu Özmercan, E. ve Kemer, B. (2024). Ölçek geliştirme konulu tezlerin ölçek geliştirme standartlarına göre incelenmesi. *Marmara Üniversitesi Atatürk Eğitim Fakültesi Eğitim Bilimleri Dergisi, 60*(60), 190-206. DOI: 10.15285/maruaebd.1486846

Ergene, Ö. (2020). Matematik eğitimi alanında ölçek geliştirme ve ölçek uyarlama makaleleri: Betimsel içerik analizi. *Yaşadıkça Eğitim*, *34*(2), 360-383. https://doi.org/10.33308/26674874.2020342207

Erkuş, A. (2012). *Psikolojide ölçme ve ölçek geliştirme*. Pegem Akademi.

Erkuş, A. (2017). Ölçek geliştirme ve uyarlama çalışmalarındaki sorunlar ile yazım ve değerlendirilmesi. *Eğitim Bilimlerinde Yenilikler ve Nitelik Arayışı*, 1211–1224. doi: 10.14527/9786053183563.075

Erkuş, A. (2019). *Psikolojide ölçme ve ölçek geliştirme-I: Temel kavramlar ve işlemler. (4. Baskı).* Pegem Akademi.

Erol, A., Yurdakal, H. İ. & Tekin Karagöz, C. (2023). Metaverse/meta-education belief scale. *Malaysian Online Journal of Educational Technology,* 11(2), 94-107. http://dx.doi.org/10.52380/mojet.2023.11.2.461

Eyüp, B., & Kayhan, S. (2023). Pre-service turkish language teachers' anxiety and attitudes toward artificial intelligence. *International Journal of Education and Literacy Studies*, *11*(4), 43-56. https://doi.org/10.7575/aiac.ijels.v.11n.4p.43

Ferikoğlu, D. & Akgün, E. (2022). An investigation of teachers' artificial intelligence awareness: a scale development study. *Malaysian Online Journal of Educational Technology,* 10(3), 215-231. https://doi.org/10.52380/mojet.2022.10.3.407

Frey, C. B. & Osborne, M. (2013). *The future of employment*. Published by the Oxford Martin Programme on Technology and Employment. https://sep4u.gr/wp-content/uploads/The_Future_of_Employment_ox_2013.pdf

Guan, C., Mou, J. & Jiang, Z. (2020). Artificial intelligence innovation in education: a twenty-year data-driven historical analysis. *International Journal of Innovation Studies*, 4(4), 134-147. https://doi.org/10.1016/j.ijis.2020.09.001

Gül, Ş., & Sözbilir, M. (2015). Fen ve matematik eğitimi alanında gerçekleştirilen ölçek geliştirme araştırmalarına yönelik tematik içerik analizi. *Eğitim ve Bilim*, *40*(178), 85-102. http://dx.doi.org/10.15390/EB.2015.4070

Hair, J. F., Black, W. C., Babin, B. J. & Anderson, R. E. (2014). *Multivariate data analysis (7th ed.).* Pearson.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory (Vol. 2)*. Sage.

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning*. Center for Curriculum Redesign.

Hwang, H. S., Zhu, L. C. & Cui, Q. (2023). Development and validation of a digital literacy scale in the artificial intelligence era for college students. *KSII Transactions on Internet and Information Systems (TIIS)*, *17*(8), 2241-2258. https://doi.org/10.3837/tiis.2023.08.016

Jang, Y., Choi, S. & Kim, H. (2022). Development and validation of an instrument to measure undergraduate students' attitudes toward the ethics of artificial intelligence (AT-EAI) and analysis of its difference by gender and experience of AI education. *Educ Inf Technol* 27, 11635–11667. https://doi.org/10.1007/s10639-022-11086-5

Kahraman, A., Aytekin, M. Ş., & Alparslan, Ö. (2021). Türkiye'de emzirme ile ilgili ölçeklerin ölçek uyarlama adımlarının incelenmesi. *Cumhuriyet Üniversitesi Sağlık Bilimleri Enstitüsü Dergisi*, *6*(3), 173-180. https://doi.org/10.51754/cusbed.901407

Karaca, O., Çalışkan, S. A. & Demir, K. (2021). Medical artificial intelligence readiness scale for medical students (MAIRS-MS) – Development, validity and reliability study. *BMC Med Educ* 21, 1-9. https://doi.org/10.1186/s12909-021-02546-6

Karaoğlan Yılmaz, F. G. & Yılmaz, R. (2023). Yapay zekâ okuryazarlığı ölçeğinin Türkçeye uyarlanması. *Bilgi ve İletişim Teknolojileri Dergisi*, 5(2), 172-190. https://doi.org/10.53694/bited.1376831

Kaya, F., Aydin, F., Schepman, A., Rodway, P., Yetişensoy, O. & Demir Kaya, M. (2024). The roles of personality traits, ai anxiety, and demographic factors in attitudes toward artificial intelligence, *International Journal of Human–Computer Interaction,* 40(2), 497-514. https://doi.org/10.1080/10447318.2022.2151730

Kılıç, A. F. (2023). (Ed.). *R programlama diliyle A'dan Z'ye ölçek geliştirme*. Nobel Akademik Yayıncılık.

Kılıç, A. F., & Uysal, İ. (2022). To what extent are item discrimination values realistic? A new index for two-dimensional structures. *International Journal of Assessment Tools in Education*, 9(3), 728-740. https://doi.org/10.21449/ijate.1098757

Koğar, H. (2021). *R ile geçerlik ve güvenirlik analizleri.* Pegem Akademi.

Konu Kadirhanoğulları, M. (2024). Biyoloji eğitiminde gerçekleştirilen ölçek geliştirme araştırmalarına bir bakış. *Yaşadıkça Eğitim, 38*(1), 184–198. https://doi.org/10.33308/26674874.2024381544

Laupichler, M. C., Aster, A., Haverkamp, N. & Raupach, T. (2023). Development of the "scale for the assessment of non-experts' ai literacy" –An exploratory factor analysis. *Computers in Human Behavior Reports*, *12*, 100338. https://doi.org/10.1016/j.chbr.2023.100338

Laupichler, M. C., Aster, A., Perschewski, J.-O. & Schleiss, J. (2023). Evaluating ai courses: A valid and reliable instrument for assessing artificial-intelligence learning through comparative self-assessment. *Educ. Sci*., 13(10), 978. https://doi.org/10.3390/educsci13100978

Long, D., Magerko, B. (2020). What is AI literacy? Competencies and design considerations. *In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems,* 1–16. Honolulu, HI, USA, https://doi.org/10.1145/3313831.3376727

Lord, F. M. (2012). *Applications of item response theory to practical testing problems*. Routledge.

Luckin, R. (2017). *Towards Artificial Intelligence-based Assessment Systems*. Pearson Education.

Morales-García, W.C., Sairitupa-Sanchez, L.Z., Morales-García, S.B. & Morales-García, M. (2024). Adaptation and psychometric properties of a brief version of the general self-efficacy scale for use with artificial intelligence (GSE-6AI) Among University Students. *In Frontiers in Education* (Vol. 9, p. 1293437). Frontiers Media SA. https://doi.org/10.3389/feduc.2024.1293437

Morales-García, W. C., Sairitupa-Sanchez, L. Z., Morales-García, S. B. & Morales-García, M. (2024). Development and validation of a scale for dependence on artificial intelligence in university students. In *Frontiers in Education.* 9:1323898. Frontiers Media SA. https://doi.org/10.3389/feduc.2024.1323898

Nazaretsky, T., Cukurova, M. & Alexandron, G. (2022). An instrument for measuring teachers' trust in aı-based educational technology. In *LAK22: 12th international learning analytics and knowledge conference,* 56-66. https://doi.org/10.1145/3506860.3506866

Ng, D. T. K., Wu, W., Leung, J. K. L., Chiu, T. K. F., & Chu, S. K. W. (2024). Design and validation of the AI literacy questionnaire: The affective, behavioural, cognitive and ethical approach. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.13411

Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric theory (3rd ed.).* McGraw-Hill.

Olgun, G. & Alatlı, B. (2021). Türkiye'de ergenlere yönelik ölçek geliştirme ve uyarlama çalışmalarının incelenmesi. *TEBD*, *19*(1), 568-592. https://doi.org/10.37217/tebd.849954

Öztemel, E. (2018). Eğitimde yeni yönelimlerin değerlendirilmesi ve eğitim 4.0. *Üniversite Araştırmaları Dergisi*, *1*(1), 25-30. Doi: 10.26701/uad.371662

Pedro, F., Subosa, M., Rivas, A. & Valverde, P. (2019). *Artificial intelligence in education: challenges and opportunities for sustainable development.* United Nations Educational, Scientific and Cultural Organization (UNESCO).

Polat, S., & Ay, O. (2016). Meta-sentez: Kavramsal bir çözümleme. *Eğitimde Nitel Araştırmalar Dergisi*, 4(2), 52-64. https://doi.org/10.14689/issn.2148-2624.1.4c2s3m

Polatgil, M. & Güler, A. (2023). Yapay zekâ okuryazarlığı ölçeğinin Türkçeye uyarlanması. *Sosyal Bilimlerde Nicel Araştırmalar Dergisi*, 3(2), 99-114.

Roll, I. & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence in Education 26*(2), 582–599. Doi:10.1007/s40593-016-0110-3

Savaş, H., & Pazar, B. (2021). Cerrahi hastalıkları hemşireliği lisansüstü tezlerinde geliştirilen ölçeklerin analizi. *Türk Hemşireler Derneği Dergisi*, 2(2), 111-124.

Schepman, A., & Rodway, P. (2023). The general attitudes towards artificial intelligence scale (GAAIS): confirmatory validation and associations with personality, corporate distrust, and general trust. *International Journal of Human–Computer Interaction*, *39*(13), 2724–2741. https://doi.org/10.1080/10447318.2022.2085400

Suh, W. & Ahn, S. (2022). Development and validation of a scale measuring student attitudes toward artificial intelligence. *Sage Open*, 12(2), 1-12. https://doi.org/10.1177/21582440221100463

Şahin, M. G., & Boztunç Öztürk, N. (2018). Eğitim alanında ölçek geliştirme süreci: Bir içerik analizi çalışması. *Kastamonu Eğitim Dergisi, 26*(1), 191-199. doi:10.24106/kefdergi.375863

Tan, Ş. (2020). *Öğretimde ölçme ve değerlendirme. (14. Baskı).* Pegem Akademi.

Terzi, R. (2020). An adaptation of artificial ıntelligence anxiety scale into turkish: Reliability and validity study. *International Online Journal of Education and Teaching*, 7(4), 1501-1515. http://iojet.org/index.php/IOJET/article/view/1031

Timms, M. J. (2016). Letting artificial intelligence in education out of the box: Educational cobots and smart classrooms. *Int J Artif Intell Educ* 26, 701–712. https://doi.org/10.1007/s40593-016-0095-y

Wang, B., Rau, P. L. P. & Yuan, T. (2023). Measuring user competence in using artificial intelligence: validity and reliability of artificial intelligence literacy scale. *Behaviour & information technology*, *42*(9), 1324-1337. https://doi.org/10.1080/0144929X.2022.2072768

Wang, Y. Y. & Wang, Y. S. (2019). Development and validation of an artificial intelligence anxiety scale: an initial application in predicting motivated learning behavior, Interactive Learning Environments, DOI:10.1080/10494820.2019.1674887

Wang, Y. Y. & Chuang, Y. W. (2024). Artificial ıntelligence self-efficacy: Scale development and validation. *Education and Information Technologies*, 29(4), 4785-4808. https://doi.org/10.1007/s10639-023-12015-w

Xie, X. & Wang, T. (2024). Artificial intelligence: A help or threat to contemporary education. should students be forced to think and do their tasks independently?. *Educ Inf Technol* 29, 3097–3111. https://doi.org/10.1007/s10639-023-11947-7

Yavuz Aksakal, N. & Ülgen, B. (2021). Artificial ıntelligence and jobs of the future. *TRT Akademi*, 6(13), 834-852. https://doi.org/10.37679/trta.969285

Yıldırım, A. & Şimşek, H. (2021). *Sosyal bilimlerde nitel araştırma yöntemleri* (12. baskı). Pegem Akademi.

Yılmaz, A., & Özgül, İ. (2023). An examination of scale development studies in the field of educational sciences within the scope of international standards. *International Journal of Eurasian Education and Culture*, 8(24), 2895-2920. http://dx.doi.org/10.35826/ijoecc.1810

Yilmaz, F. G. K., Yilmaz, R., & Ceylan, M. (2023). Generative artificial ıntelligence acceptance scale: A validity and reliability study. *International Journal of Human–Computer Interaction*, 1-13. https://doi.org/10.1080/10447318.2023.2288730

Yurdugül, H. (2005, 28–30 Eylül). Ölçek geliştirme çalışmalarında kapsam geçerliği için kapsam geçerlik indekslerinin kullanılması. *XIV. Ulusal Eğitim Bilimleri Kongresi*, Pamukkale Üniversitesi Eğitim Fakültesi, 1, 771-774. Denizli, Türkiye.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education. *International Journal of Educational Technology in Higher Education*, 16(1), 39.

EXTENDED ABSTRACT

1. INTRODUCTION

In the introduction of the study, the increasing use of artificial intelligence (AI) technologies in education and the importance of scales developed to measure the effects of this use were emphasized. Understanding and assessing the impacts of AI in education is becoming increasingly important for teachers, students and other stakeholders. The need to develop valid and reliable scales that measure current attitudes, concerns, perceptions, and literacy levels regarding the adoption of AI technologies in education has emerged. The purpose of this study is to examine the validity and reliability studies of the scales developed for the use of artificial intelligence in education. The study aims to evaluate the general characteristics of the scales developed in the field of education related to artificial intelligence and their suitability in validity and reliability studies. In this direction, the contents, methods used and study groups of existing scales for the use of AI in education were analyzed in detail.

2. METHOD

In the methodology section of the study, it was stated that the research was designed as a qualitative study and content analysis method was used. The basic data of the study consisted of 24 different scales developed or adapted for the use of AI in education. These scales include scales adapted from foreign languages and originally developed scales. In selecting the scales, certain criteria were taken into consideration, such as the availability of validity and reliability analyses of the scales, sample size and the types of validity used. Within the scope of content analysis, the scales were classified and evaluated in terms of their development purposes, validity and reliability analyses, sample sizes, study groups and other methodological details. Factors such as the most preferred types of participants as study groups, the effect of sample size on the validity of the scales, and the validity analyses used were examined in detail. The data were categorized using qualitative analysis techniques and summarized with descriptive statistics.

3. FINDINGS, DISCUSSION AND RESULTS

In the discussion and conclusion section, the general characteristics of the scale development studies on the use of artificial intelligence in education and the findings that are compatible with the literature are discussed. Firstly, it was stated that approximately one third of the scales were adapted from foreign languages into Turkish, while the rest were original scales. It was observed that the majority of the scales were developed on AI literacy and attitude. It was emphasized that there are few studies especially in the field of AI literacy and there is a need for more research in this field (Cin Şeker & Yücel Çetin, 2022; Konu Kadirhanoğulları, 2024). The widespread preference of attitude scales is a common trend in the literature (Ergene, 2020). However, it was concluded that issues such as anxiety, perception and self-efficacy are also important and more scales should be developed on these issues.

Construct validity was the most commonly used validity type during the development of the scales. With the use of computer technologies in the analyses, the use of Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) in construct validity analyses increased. Nearly half of the scales analyzed used only CFA, and in one third, both EFA and CFA were applied together. The fact that different samples were not used for the analyses in one fourth of the studies in which EFA and CFA were used together shows that validity analyses contain some methodological deficiencies. In the literature, it is generally recommended to use EFA for construct validity, followed by CFA with a different sample (Cabrera-Nguyen, 2010). However, it was observed that this recommendation was followed only in almost half of the studies examined.

It was observed that surface validity analyses were also common, but item-total correlation for content validity was not sufficiently examined. This situation leads to confusion between content validity and face validity in the literature. It is a common situation in the literature that surface validity analyses conducted by applying expert opinions are confused with content validity (Ergene, 2020; Gül & Sözbilir, 2015). However, the limited use of test-retest analyses is considered as an incomplete approach in terms of determining the stability of the scales.

It was found that Cronbach's Alpha coefficient was mostly used in reliability analyses and this coefficient was generally above 0.90. The fact that Cronbach's Alpha is the most widely used internal consistency criterion in scale development studies stands out as a finding consistent with the literature (Gül & Sözbilir, 2015). However, the low use of the test-retest method is attributed to the tendency of researchers to avoid wasting time and resources. This is a finding that is frequently criticized in the literature (Ayyıldız, 2022; Gül & Sözbilir, 2015).

In conclusion, this study on the scales developed for the use of artificial intelligence in education shows that the validity and reliability analyses are generally consistent with the literature. However, deficiencies such as the scarcity of scales especially for K12 students, inadequacy of discriminant validity analyses and low use of test-retest method indicate that research in the field should be improved. Accordingly, it is suggested that more methodologically comprehensive approaches should be adopted in future scale development studies.

This study provides important contributions to understand the current status of scale development processes for the use of artificial intelligence in education and to identify the deficiencies in this field. It is concluded that studies on the development of valid and reliable scales should be increased for the effective evaluation of artificial intelligence technologies in education.

ARAŞTIRMANIN ETİK İZNİ

Bu çalışmada "Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesi" kapsamında uyulması gerektiği belirtilen tüm kurallara uyulmuştur.

Etik kurul izin bilgileri

Bu çalışma anket, mülakat, odak grup çalışması, gözlem, deney, görüşme teknikleri kullanılarak katılımcılardan veri toplanmasını gerektiren nitel ya da nicel yaklaşımlarla yürütülen, insan ve hayvanların (materyal/veriler dâhil) deneysel ya da diğer bilimsel amaçlarla kullanılan ve kişisel verilerin korunması kanunu gereğince retrospektif çalışmalar kapsamına dahil olmadığı için etik kurul izni gerektirmemektedir

ARAŞTIRMACILARIN KATKI ORANI

Araştırmacılar, araştırmaya eşit oranda katkı sağlamışlardır. Bu doğrultuda her bir yazarın katkı oranı %50'dir. Araştırmacıların araştırmanın hangi aşamalarına katkıda bulundukları aşağıda açık bir şekilde ifade edilmiştir.

Yazar 1: Giriş, verilerin toplanması, analizi ve bulguların yazımı

Yazar 2: Yöntem, sonuç, tartışma ile danışmanlık ve redaksiyon.

ÇATIŞMA BEYANI

Araştırmada herhangi bir çıkar çatışması yoktur.