



YABANCI DİLDE KONUŞMA BECERİSİNİN ÜRETKEN YAPAY ZEKÂ İLE DEĞERLENDİRİLMESİ

EVALUATING FOREIGN LANGUAGE SPEAKING PROFICIENCY USING GENERATIVE ARTIFICIAL INTELLIGENCE

Oğuz İbrahim DÜNDAR

Anadolu Üniversitesi, Yabancı Diller Yüksekokulu

oidundar@anadolu.edu.tr

ORCID: 0000-0002-9574-6125

ÖZ

ABSTRACT

Geliş Tarihi:

19.07.2025

Kabul Tarihi:

25.10.2025

Yayın Tarihi:

04.12.2025

Anahtar Kelimeler

Hızlı Derleme,
Konuşma Becerisi,
Ölçme-
Değerlendirme,
Üretken Yapay Zekâ,
Yapay Zekâ.

Keywords

Artificial
Intelligence,
Assessment,
Generative Artificial
Intelligence,
Rapid Review,
Speaking
Proficiency.

Bu hızlı derleme çalışması, üretken yapay zekâ (ÜYZ) destekli araçların yabancı dil öğrenen bireylerin konuşma becerilerinin değerlendirilmesindeki rolünü tematik analiz yöntemiyle incelemektedir. Çalışma kapsamında, alan yazında konuya odaklanan 67 çalışma taranmış; 2018-2025 yılları arasında yayımlanan, ÜYZ araçlarının kullanıldığı ve yabancı dil öğrenenlerle yürütülmüş 7 ampirik araştırma belirlenen ölçütlere göre derinlemesine analiz edilmek üzere seçilmiştir. Veri toplama sürecinde Scopus, Google Scholar, Web of Science ve EBSCO gibi dört büyük akademik veri tabanı kullanılmıştır. Elde edilen bulgular dört ana tema etrafında toplanmaktadır: (1) otomatik geribildirim ve bireyselleştirilmiş değerlendirme (2) dil yeterlilikleri ve öğrenen özerkliği üzerindeki etkiler, (3) yenilikçi teknik yaklaşımlar, (4) sistemsel sınırlılıklar ve etik meseleler. Büyük dil modelleri ve otomatik konuşma tanıma teknolojileri, öğrencilere anlık, detaylı ve hedefe yönelik geribildirim sunarak öğrenme süreçlerini bireyselleştirmekte ve öğretmenlerin ölçme-değerlendirme yükünü önemli ölçüde azaltmaktadır. Ancak bu sistemler, kültürel duyarlılık, duygusal zekâ ve bağlamsal esneklik açısından hâlâ çeşitli sınırlılıklar taşımaktadır. Çalışma, ÜYZ tabanlı değerlendirme uygulamalarının pedagojik hedeflerle uyumlu, etik ve kültürel açıdan duyarlı bir yaklaşımla ele alınmasının gerekliliğini vurgulamaktadır.

This rapid review study explores the role of generative artificial intelligence (GAI)-powered tools in assessing the speaking skills of foreign language learners, employing a thematic analysis approach. A total of 67 studies focused on the topic were initially identified through a literature search. Of these, seven empirical studies, published between 2018 and 2025, involving foreign language learners, and utilizing GAI tools, were selected for in-depth analysis based on defined inclusion criteria. Data collection was carried out using four major academic databases: Scopus, Google Scholar, Web of Science, and EBSCO. The findings are organized under four main themes: (1) automated feedback and personalized assessment (2) impacts on linguistic competence and learner autonomy (3) innovative technical approaches, and (4) systemic limitations and ethical issues. Large language models and automatic speech recognition technologies provide learners with immediate, detailed, and targeted feedback, thereby personalizing learning and significantly reducing teachers' assessment workload. However, these systems still face limitations in terms of cultural sensitivity, emotional intelligence, and contextual adaptability. The study emphasizes the importance of adopting GAI-based assessment practices that align with pedagogical goals and are ethically and culturally responsive.

DOI: <https://doi.org/10.30783/nevsosbilen.1746385>

Atıf/Cite as: Dündar, O. I. (2025). Yabancı dilde konuşma becerisinin üretken yapay zekâ ile değerlendirilmesi. *Nevşehir Hacı Bektaş Veli Üniversitesi SBE Dergisi, Sosyal Bilimlerde Yapay Zekâ: Kuram, Uygulama ve Gelecek Perspektifleri Özel Sayısı*, 169-184.

Giriş

Son yıllarda yapay zekâ (YZ), özellikle üretken yapay zekâ (ÜYZ) teknolojilerinin hızla gelişmesi, eğitim alanında köklü değişimlerin kapısını aralamıştır. Dil öğrenimi süreçlerinde, öğrencilerin ihtiyaçlarına bireysel düzeyde yanıt verebilen YZ tabanlı araçlar, geleneksel öğretim yöntemlerine yenilikçi alternatifler sunmaktadır. Bu teknolojiler, özellikle sözlü dil üretiminin değerlendirilmesinde yaşanan zaman, kaynak ve objektiflik sorunlarını hafifletme potansiyeli taşımaktadır (Zaghlool & Khasawneh, 2023). Bu bağlamda, ÜYZ sistemlerinin, konuşma becerilerinin öğretimi ve değerlendirilmesinde yeni eğilimlerin ortaya çıkmasına zemin hazırladığı ifade edilebilir. Özellikle ChatGPT gibi büyük dil modelleri, kişiselleştirilmiş etkileşim, sürekli geri bildirim ve konuşma pratiği olanakları sunarak, öğrencilerin konuşma becerilerini çeşitli boyutlarda desteklemektedir (Jeon vd., 2023). Alan yazındaki bulgular, bu tür araçların sadece dil yeterliliklerini değil, aynı zamanda öğrencilerin konuşma öz yeterliği, iletişim istekliliği ve öğrenme motivasyonunu da olumlu yönde etkilediğini göstermektedir (Xu vd., 2024; Çelik vd., 2025; Wang, 2025). Öğrencilerin tekrar yapma, hata düzeltme ve kendini ifade etme süreçlerinde daha özgür bir ortam bulmaları, bu gelişimin temel etkenleri arasında sayılmaktadır (Bae vd., 2024; Xu vd., 2024).

Bununla birlikte, bu sistemlerin pedagojik katkılarının yanı sıra sınırlılıkları da göz ardı edilmemelidir. Yapay zekâ sistemlerinin zaman zaman halüsinatif geri bildirimler üretmesi (modelin ürettiği çıktının verilen kaynak girdiye veya gerçek dünyadaki doğrulara uyumlu olmaması, yani aslı olmayan, tutarsız, anlamsız ya da uydurma içerikler üretmesi) insan etkileşiminin eksikliğiyle birlikte öğrenme sürecini olumsuz yönde etkileyebilecek önemli sınırlılıklar oluşturmaktadır (Park & Lee, 2024). Ayrıca, bazı araştırmalar, öğrencilerin yapay zekâyı aşırı güven geliştirmesinin akademik bütünlük ve eleştirel düşünme becerileri açısından potansiyel riskler taşıdığına dikkat çekmektedir (Khatri & Karki, 2023; Gammoh, 2024; Zhai vd., 2024).

Bu bağlamda, bu çalışma, ÜYZ destekli konuşma değerlendirme araçlarının konuşma becerisi bağlamında ne ölçüde etkili olduğunu anlamaya yönelik olarak yürütülmüş bir hızlı derleme çalışmasıdır. Bu çalışma, kapsamlı sistematik derlemelere kıyasla daha hızlı yanıt üretmek üzere yapılandırılmış, sınırlı ama nitelikli çalışmaları tematik olarak inceleyen bir yöntem (Dobbins, 2017) olan hızlı derleme yöntemi ile gerçekleştirilmiştir. Bu çalışmada, ÜYZ'nin dil öğrenme süreçlerine katkılarını ve sınırlılıklarını ortaya koyan yakın dönemli akademik yayınlar incelenmiş, bulgular tematik çerçevede yorumlanmıştır. Elde edilen sonuçlar, yapay zekâ teknolojilerinin pedagojik açıdan nasıl ve ne şekilde kullanılabileceği sorusuna bütüncül bir perspektif sunmayı amaçlamaktadır.

Alanyazın

YZ, insan zekâsını taklit edebilen bilgisayar sistemlerini ifade eden geniş bir kavramdır. Bu sistemler, veriler üzerinden öğrenme, akıl yürütme, karar verme ve problem çözme (Görgüt, 2024) gibi bilişsel süreçleri simüle ederek çeşitli görevleri bağımsız biçimde gerçekleştirebilir. YZ kavramı, ilk olarak 1956 yılında Dartmouth Konferansı'nda akademik düzeyde dile getirilmiş; yıllar içinde özellikle makine öğrenmesi, derin öğrenme ve doğal dil işleme gibi alt alanların gelişimiyle ciddi ilerlemeler kaydedilmiştir. Günümüzde YZ sistemleri sağlık (Schwalbe ve Wahl, 2020), hukuk (Surden, 2018), ekonomi (Bolton vd., 2018) ve özellikle eğitim (Malfatti, 2025) gibi sosyal alanlarda da yaygın biçimde kullanılmaktadır. Yapay zekânın temel bileşenleri arasında algoritmalar, büyük veri, hesaplama gücü ve etkileşimli öğrenme mekanizmaları (Lu, 2019) yer almaktadır. Bu teknolojiler, statik bilgi sunmanın ötesine geçerek kullanıcılarla aktif ve kişiselleştirilmiş etkileşimler kurabilmektedir (Najafov, 2024). Söz konusu nitelikler, eğitim gibi bireysel ihtiyaçların ön planda olduğu alanlarda YZ'nin potansiyelini artırmaktadır.

Yapay zekâ alanında iki temel yaklaşımın bulunduğu unutulmamalıdır: ilki, kurallara dayalı karar ağaçları ve istatistiksel modeller gibi geleneksel algoritmalar; ikincisi ise ÜYZ olarak bilinen, büyük veri kümeleri üzerinden öğrenerek yeni içerikler üretebilen derin öğrenme tabanlı sistemlerdir (Dündar, 2024). Bu iki yaklaşım arasındaki temel fark, ilkinde bilgi işleme süreci önceden tanımlı kurallara dayanırken, ÜYZ sistemlerinin örüntüleri verilerden öğrenerek kendi çıktısını oluşturabilmesidir.

Son yıllarda dikkat çeken teknolojik gelişmeler arasında yer alan ÜYZ, yalnızca veri analizi yapmakla kalmayıp, yeni ve özgün içerikler de üretebilme kapasitesiyle öne çıkmaktadır. ÜYZ, büyük dil modelleri aracılığıyla yazılı, sözlü, görsel ve hatta çoklu modal veriler oluşturabilmektedir (Kumar vd., 2023). Özellikle GPT (Generative Pre-trained Transformer) mimarisine dayanan modeller, insan benzeri diyaloglar kurabilmekte, karmaşık dil yapılarını çözümlayebilmekte ve yaratıcı içerikler (Yenduri vd., 2023) üretebilmektedir. ÜYZ'nin eğitimdeki

etkisi, geleneksel öğrenme yaklaşımlarının ötesine geçen yeni pedagojik olanakları beraberinde getirmiştir. Öğrencilere sürekli geri bildirim sağlama, bireysel öğrenme yolları oluşturma ve öğrenme sürecini özelleştirme gibi işlevler (Sreen & Majid, 2024) bu teknolojinin öne çıkan özellikleri arasında yer almaktadır. ChatGPT gibi modeller, yalnızca bilgi sunmakla kalmayıp, kullanıcıyla etkileşim kurarak (Mema vd., 2024) süreci dinamik bir öğrenme deneyimine dönüştürebilmektedir. YZ tabanlı sistemlerin eğitim alanında kullanımı, öğrenme süreçlerinin hem bireysel hem de toplu düzeyde iyileştirilmesine yönelik güçlü bir potansiyel taşımaktadır. Otomatik notlandırma sistemlerinden (Vetrivel vd., 2025) öğrenme analitiklerine (Salas-Pilco vd., 2022) sanal asistanlardan (Page & Gehlbach, 2017) kişiselleştirilmiş öğretim yazılımlarına (Hashim vd., 2022) kadar pek çok uygulama alanı mevcuttur. ÜYZ sistemlerinin öne çıkan avantajları arasında ölçeklenebilirlik, objektif değerlendirme, anlık geri bildirim, erişilebilirlik ve bireyselleştirilmiş öğrenme ortamları oluşturma kapasitesi (Ayenı vd., 2024; Mahmoud & Sørensen, 2024) yer almaktadır.

Bu bağlamda, öğrencilerin farklı hız, ilgi ve yetenek düzeylerine göre uyarlanabilir içerik sunabilen yapay zekâ destekli platformlar, geleneksel sınıf temelli eğitimin sınırlarını genişletmektedir (Dündar & Aydoğu, 2024a). Ayrıca, öğretmenlerin değerlendirme ve planlama yükünü hafifletmesi sayesinde daha fazla etkileşim ve rehberliğe zaman ayrılması mümkün hâle gelmektedir. Bununla birlikte, pedagojik uyumluluk (Murtiningsih vd., 2024), etik sorumluluklar (Stahl & Eke, 2024), veri güvenliği (Derner & Batistič, 2023) ve akademik dürüstlük (Sullivan vd., 2023) gibi konular yapay zekânın eğitimdeki yaygınlaşmasıyla birlikte tartışma konusu olmaktadır. Bu nedenle YZ'nin öğrenme ortamlarına entegrasyonu, yalnızca teknolojik değil; aynı zamanda pedagojik ve etik ilkeler çerçevesinde yapılandırılmalıdır.

Konuşma becerileri, yabancı dil öğretiminin temel bileşenlerinden biridir ve dilin sosyal bir araç olarak kullanılabilmesini mümkün kılar. Zira,

yabancı dil öğrenenlerin, konuşma becerisinde başarılı sayılabilmeleri için temel gereksinimlerinden biri Avrupa ortak dil çerçeve metninde de yer aldığı üzere sözlü üretim becerisi kazanabilmeleri; bir diğeri ise rahat ve akıcı bir biçimde diyalog kurabilmeleri, bir başka ifade ile öğrendiği dilin toplumuna ait iletişim ortamlarında sosyal aktör olarak aktif bir biçimde yer alabilmeleridir. (İnal, 2020 s.189)

Bu beceri, yalnızca doğru telaffuzdan ibaret olmayıp; akıcılık, sözdizimi uygunluğu, pragmatik yeterlik, telaffuz, vurgu ve beden dili gibi (Neupane, 2025) çok yönlü alt becerileri içermektedir. Geleneksel dil öğretiminde konuşma becerileri, genellikle bire bir diyaloglar, grup çalışmaları, rol yapma etkinlikleri ve sınıf içi sunumlarla desteklenmeye çalışılmıştır. Ancak bu uygulamalar, büyük sınıflarda her öğrenciye yeterli süre ve geri bildirim sağlama açısından çeşitli sınırlılıklar (Dündar, 2024) barındırmaktadır. Günümüzde, özellikle dijital teknolojilerin gelişmesiyle birlikte konuşma becerilerinin öğretimi daha esnek ve teknolojik çözümlerle desteklenebilir hâle gelmiştir. Bu noktada yapay zekâ, bireyselleştirilmiş geri bildirimler (Xu vd., 2023), ses analizi (Terzopoulos & Satratzemi, 2019), otomatik telaffuz düzeltmeleri (Dai, 2022) ve yapay partnerlerle konuşma simülasyonları (Dündar & Aydoğu, 2024b) gibi araçlarla öğretim süreçlerini dönüştürmektedir. YZ'nin konuşma becerileri öğretiminde kullanımı, özellikle otomatik konuşma tanıma (OKT-ASR), doğal dil işleme (DDİ-NLP) ve üretken modellerin gelişmesiyle ivme kazanmıştır. ÜYZ, öğrencilere anlık konuşma pratiği sunma, hatalarını fark etme ve düzeltme, bağlamsal uygunluk geliştirme gibi alanlarda destek sağlamaktadır. Ayrıca, öğrencilerin hata yapma kaygısını azaltması (Hawanti & Zubaydullovna, 2023), alıştırmaya yapma olanakları sunması (Zou vd., 2024) ve bireysel öğrenme yollarına göre şekillendirilmiş deneyimler sunabilmesi (Tapalova & Zhiyenbayeva, 2022), bu araçların dikkat çekici yönlerindendir. ÜYZ sistemlerinin bir diğer avantajı ise, ölçme ve değerlendirme süreçlerine getirdiği katkılardır. Otomatik puanlama, bireysel performans izleme, gelişim takibi gibi uygulamalar, öğretmenlerin yükünü azaltmaktadır (Hooda vd., 2022; Halagatti vd., 2023; Owan vd., 2023). Bu gelişmeler, yapay zekâ tabanlı teknolojilerin yalnızca destekleyici araçlar değil, aynı zamanda konuşma becerilerinin öğretimi ve değerlendirilmesinde dönüştürücü bir role sahip olabileceğini ortaya koymakta; bu nedenle alan yazında bu araçların etkilerine yönelik analizlere duyulan ihtiyacı artırmaktadır.

Yöntem

Bu çalışma, ÜYZ destekli araçların yabancı dil öğrenen bireylerin konuşma becerileri üzerindeki etkilerini anlamaya yönelik olarak gerçekleştirilmiş bir hızlı derleme çalışmasıdır. Bu yöntem, daha geniş kapsamlı sistematik derlemelere kıyasla daha az sayıda güncel çalışmayı daha kısa sürede analiz etme olanağı sunar (Dobbins, 2017). Bu çalışmada sistematik derleme yerine hızlı derleme yöntemi tercih edilmesinin temel

nedenleri arasında, konuyla ilgili henüz sınırlı sayıda ampirik ve nitelikli çalışmanın mevcut olması, derinlemesine sistematik analiz için yeterli veri havuzunun bulunmaması ve alanın hâlen gelişmekte olması yer almaktadır. Ayrıca, ÜYZ araçlarının konuşma becerisinin değerlendirilmesi için kullanımı oldukça güncel bir konu olduğundan, hızlı ve pratik bir şekilde mevcut alan yazını analiz ederek eğitimciler ve araştırmacılar için öncül bir bakış açısı sunmak amaçlanmıştır. Konunun yeniliği nedeniyle, ampirik veri sunan preprint ve hakemli konferans bildirilerine de sınırlı sayıda yer verilmiştir. Bu bağlamda mevcut çalışmada, konuşma becerilerinin, ÜYZ aracılığıyla değerlendirmesini inceleyen güncel çalışmalar belirlenmiş ve içerik açısından tematik olarak analiz edilmiştir. Çalışmaların seçimi, açık ve şeffaf kriterlere dayandırılmış; analiz süreci yapılandırılmış bir tematik okuma çerçevesinde yürütülmüştür.

Veri Toplama Süreci

Alan yazın taraması, dil modellerinin yaygınlaşmaya başladığı Haziran 2018 ile Haziran 2025 tarihleri arasında yayımlanmış, ÜYZ sistemlerinin konuşma becerileri üzerindeki etkilerini inceleyen özgün çalışmalara odaklanarak yürütülmüştür. Veri toplama süreci için dört büyük akademik veri tabanı kullanılmıştır: Scopus, Google Scholar, Web of Science ve EBSCO. Aşağıdaki anahtar kelimelerle arama yapılmıştır:

“ChatGPT AND speaking assessment”, “Generative artificial intelligence AND speaking assessment” “artificial intelligence AND speaking assessment”

Tarama sonucunda, alan yazında konuya odaklanan 67 çalışma ön elemeye alınmış, bunlar arasından belirlenen dâhil etme ölçütlerine uyan 7 çalışma derinlemesine analiz için seçilmiştir. Çalışmanın kapsamı, ÜYZ uygulamalarının konuşma becerisinin değerlendirilmesine etkileriyle sınırlıdır. Bu nedenle “language assessment” ve benzeri başlık altındaki çalışmalar, konuşma becerisi ile ilgili bir çıktı raporlamıyorsa kapsam dışında bırakılmış; beceri odakları karışık olan çalışmalarda ise konuşma çıktısının açıkça raporlanması dâhil edilme için önkoşul olarak aranmıştır. Çalışma kapsamında, analiz edilecek çalışmaların bilimsel geçerlilik, bağlamsal uygunluk ve metodolojik yeterlilik açısından belli ölçütleri karşılaması amaçlanmıştır. Aşağıdaki tablo, çalışmaya dâhil edilen ve hariç tutulan kaynakların belirlenmesinde kullanılan esas ölçütler özetlemektedir.

Tablo 1. Dâhil Edilme ve Hariç Tutulma Kriterleri

Dâhil Edilme Kriterleri	Hariç Tutulma Kriterleri
2018-2025 yılları arasında yayımlanmış çalışmalar.	Yalnızca okuma, yazma ya da dinleme gibi diğer dil becerilerine odaklanan çalışmalar.
Yabancı dilde konuşma becerilerini inceleyen çalışmalar.	Katılımcıları öğrenen bireylerden (her yaş ve düzey) oluşan çalışmalar.
Katılımcıları dil öğrenen bireylerden (her yaş ve düzey) oluşan çalışmalar.	Katılımcıları öğretmenler, uzmanlar veya ana dili konuşurları olan çalışmalar.
Üretken yapay zekâ araçlarının kullanıldığı çalışmalar.	Üretken yapay zekâ araçlarının kullanılmadığı çalışmalar.
Ampirik yöntemle yürütülmüş (nitel, nicel veya karma) bir araştırma olması.	Derleme, sistematik inceleme, kuramsal metin, görüş yazısı, editöryal türündeki yayınlar.

Veri Analizi

Seçilen yedi çalışmanın özet, yöntem, bulgu, tartışma ve sonuç bölümleri sistematik bir biçimde incelenmiş ve içerikleri tematik analiz yaklaşımıyla yorumlanmıştır. Tematik analiz, veri içinde yinelenen anlam örüntülerinin sistematik olarak belirlenmesi, kümelenmesi ve raporlanması süreci olarak işletilmiştir. Çalışmalar çoklu okumalarla incelenmiş, tümevarımsal açık kodlama ile veriden türeyen göstergeler çıkarılmış, benzer içerikli kodlar aday temalara gruplanmış ve tema sınırları geriye dönük denetim ile netleştirilmiştir. Otomatik geribildirim ve bireyselleştirilmiş değerlendirme temasında kodlama döngülerinde “anlık geri bildirim”, “kişiselleştirme”, “dönüt”, “iş yükü” ve “tutarlılık” gibi kodların yüksek yakınlıkla kümelenildiği; Dil yeterliliği ve öğrenen özerkliği temasında “akıcılık/doğruluk artışı”, “sözcük kapsamı gelişimi”, “öz yeterlik”, “motivasyon” ve “kaygı” kod kümelerinin öne çıktığı gözlenmiştir. Benzer biçimde, Yenilikçi teknik yaklaşımlar teması “speech LLM”, “ASR-CLM”, “çok modlu analiz” ve “genelleme” kodlarıyla; Sistemsel sınırlılıklar ve etik meseleler teması ise “halüsinasyon”, “önyargı”, “veri gizliliği” ve “akademik dürüstlük” kodlarıyla temsil edilmiştir.

Bulgular

Bu hızlı derleme kapsamında analiz edilen yedi çalışma, ÜYZ araçlarının yabancı dil öğrenen bireylerin konuşma becerilerinin değerlendirilmesine ilişkin çok boyutlu etkilerini ortaya koymaktadır.

Seçim ölçütleri doğrultusunda dâhil edilen yedi çalışmanın kimlerle, hangi araçlarla ve hangi çıktıları değerlendirdiği, tematik çözümlemenin şeffaflığını artırmak üzere Tablo 2’de sunulmuştur.

Tablo 2. İncelenen çalışmalar

Çalışma başlığı	Yazar	Sene	Yapay Zekâ aracı	Katılımcı grubu	Başlıca bulgu
Revolutionizing Language Teaching: AI in Oral Language Assessment	Gaballo, V.	2024	ÜYZ tabanlı, çok modlu (multimodal) konuşma değerlendirm e aracı	Üniversite öğrencileri, University of Padua (n=145; seviye: raporlanma mış)	Değerlendirmenin etkililiği ve verimliliği artmıştır. Puanlama süreçleri tutarlılık kazanmış ve anlık, bireyselleştirilmiş geri bildirimle desteklenmiştir. Çok modlu yaklaşım, iletişimsel yeterliğin daha nüanslı ve çok boyutlu değerlendirilmesini mümkün kılmıştır.
Research on Oral Teaching of International Chinese Language Education Assisted by ChatGPT.	Ma, L.	2024	ChatGPT	Üniversite öğrencileri, Shaanxi Normal University (n=30; deney-kontrol; seviye: Çince orta düzey)	Deney grubunda akıcılık, doğruluk ve sözcük kullanımı göstergelerinde anlamlı artış gözlenmiştir. Katılımcılar memnuniyet, motivasyon ve öz güven açısından olumlu değişim bildirmiştir. Kişiselleştirilmiş geri bildirim, öğretmen rehberliğini kolaylaştırmış ve değerlendirme iş yükünü azaltmıştır.
Assessment of L2 Oral Proficiency using Speech Large Language Models.	Ma vd.	2025	Speech LLM (Qwen2-Audio-7B-Instruct), ASR’ye gerek duymayan akış	İngilizce sınav adayları; değerlendir me setlerinde toplam n = 2,199 test başvurusu	Model, iki bağımsız test setinde konuşma puanlarını insan puanlamalarıyla yüksek düzeyde uyumlu biçimde tahmin etmiştir (Pearson korelasyon katsayısı 0,954 ve 0,938). Sınırlı veriyle eğitildiğinde ve bölüm/görev aktarımında da güçlü genelleme göstermiştir. ASR’ye ihtiyaç duymayan yaklaşım bilgi kaybını ve hesaplama maliyetini azaltmıştır.
To be with artificial intelligence in oral test or not to be: a probe into the traces of success in speaking skill, psychological well-being, autonomy, and academic buoyancy.	Sayed vd.	2024	ChatGPT	Üniversite öğrencileri, Etiyopya (n=28; üst-orta)	Gecikmeli testte konuşma becerilerinde anlamlı artış saptanmıştır. Psikolojik iyi oluş, öğrenen özerkliği düzeyleri yükselmiştir. Bireyselleştirilmiş ve anlık geri bildirim sağlandığında kaygı azalmış, motivasyon ve öz güven artmıştır.
Spoken grammar assessment using LLM.	Kopparapu vd.	2024	LLM + hibrit ASR-CLM	Yabancı dil olarak İngilizce öğrencileri, (n = 17)	Konuşma verisinden dilbilgisi değerlendirmesinin uygulanabilir olduğu gösterilmiştir. Her öğrenciye benzersiz paragraf üretilmesi ezbere dayalı hazırlığı azaltmıştır.

Human-robot interaction using ChatGPT for technology driven language learning: contextual insights from the higher education institution of Indonesia.	Muthmainnah vd.	2025	ChatGPT	Endonezya, n=7. İngiliz Dili Eğitimi yüksek lisans programına kayıtlı yabancı dil olarak İngilizce öğretmenle ri (35-42 yaş).	Katılımcılar, ChatGPT'nin girdi sunma, anlık geri bildirim üretme ve sınıf içi etkileşimi artırma yönlerinden yararlı olduğunu bildirmiştir. Bununla birlikte, insan bilişinin yerini alamaması, aşırı bağımlılık ve akademik dürüstlük ile önyargı riski gibi sınırlılıklar vurgulanmıştır. Çalışma, aracı öğretmenliğin yerine değil, etik ilkelere uygun tamamlayıcı bir araç olarak entegrasyonu önermektedir.
Exploring the Role of Artificial Intelligence (AI) Support in Assisting Students' English-Speaking Skills.	Pangestu ve Suwartono	2024	SiVi, Duolingo, ChatGPT	Endonezya, İngilizce Öğretmenli ği 6. yarıyıl öğrencileri(n=4; seviye: raporlanma mış)	Öğrenciler, telaffuz, sözcük bilgisi ve dilbilgisi alanlarında gelişme bildirmiştir. Kişiselleştirme ve otomatik değerlendirme süreçlere anlamlı katkı sağlamıştır. Bununla birlikte, sınırlı insan etkileşimi ve nüans, mizah ile karmaşık bağlamların anlaşılmasındaki güçlükler ve bazı araçların özellik/doğruluk kısıtları gözlenmiş; bu nedenle insan rehberliğinin gerekli olduğu sonucuna varılmıştır.

Derleme kapsamındaki bulgular, yapay zekâ destekli değerlendirme süreçlerinin pedagojik, teknik ve duyuşsal açılardan çeşitli katkılar sunduğunu göstermektedir. Tematik analiz neticesinde elde edilen veriler dört ana kategori altında yapılandırılmıştır: (1) otomatik geribildirim ve bireyselleştirilmiş değerlendirme (2) dil yeterliliği ve öğrenen özerkliği üzerindeki etkiler, (3) yenilikçi teknik yaklaşımlar, (4) sistemsel sınırlılıklar ve etik meseleler.

Otomatik Geribildirim ve Bireyselleştirilmiş Değerlendirme

Çalışmaların büyük çoğunluğu, ÜYZ tabanlı araçların öğrencilerin bireysel dil gelişim süreçlerine doğrudan ve çok yönlü katkı sunduğunu ortaya koymaktadır. Büyük dil modelleri temelli sistemler, öğrencilerin konuşma esnasında gerçekleştirdiği dilsel hataları anında analiz edebilmekte ve bu hatalara yönelik kişiselleştirilmiş, hedef odaklı ve açıklayıcı geribildirimler sunabilmektedir (Gaballo, 2024; Ma, 2024). Bu durum, hem öğrenme sürecinin bireysel ihtiyaçlara göre şekillenmesine olanak tanımakta hem de öğretmenlerin değerlendirme yükünü hafifletmektedir. Özellikle büyük öğrenci gruplarına sahip sınıf ortamlarında, bu tür teknolojik çözümler değerlendirme sürecinde zaman ve emek tasarrufu sağlamaktadır.

Gaballo (2024) tarafından yürütülen vaka çalışmasında, 145 kişilik bir lisans sınıfında ÜYZ destekli değerlendirme sistemlerinin kullanımıyla öğretim üyesinin ölçme süreçlerindeki iş yükü anlamlı düzeyde azalmış; öğrenciler ise hem niceliksel hem de niteliksel çok katmanlı, sürekli geribildirim alma imkânı elde etmiştir. Sistem, sadece öğrencinin doğru ya da yanlış yanıtlarını değil, aynı zamanda telaffuz, akıcılık, dil bilgisel doğruluk ve kelime kullanımı gibi alt becerileri de ölçümleyebilmiştir. Benzer biçimde, Ma (2024) tarafından orta düzeyde Çince öğrenen öğrencilerle yürütülen çalışmada, ChatGPT destekli öğretimin öğrencilerin gerçek zamanlı etkileşim kurma yetilerini geliştirdiği, kişiselleştirilmiş pratik olanağı sunduğu ve otomatik değerlendirme özelliği sayesinde öğrenme sürecini yapılandırdığı sonuçlarına ulaşmıştır. Otomatik geribildirim sistemlerinin özellikle öğrencilerin hatalarını doğrudan fark etmelerini ve bu hatalara ilişkin düzeltici geri dönüşleri anlık olarak almalarını sağlaması, geleneksel yöntemlere göre önemli bir avantaj sunmaktadır.

Sayed vd. (2024) tarafından gerçekleştirilen deneysel çalışmada ise TOEFL iBT konuşma bölümü ChatGPT aracılığıyla değerlendirilmiş; hem kısa vadede (son test) hem de uzun vadede (gecikmeli test) konuşma becerilerinde kayda değer gelişmeler gözlemlenmiştir. ChatGPT'nin sunduğu bireyselleştirilmiş geribildirimlerin, öğrenenlerin özgün eksiklerine göre yapılandırıldığı ve bu sayede kalıcı gelişim sağlandığı ifade edilmiştir. Bu

bulgu, ÜYZ tabanlı değerlendirmenin yalnızca performans takibi değil, aynı zamanda öğrenme sürecinin dinamik bir bileşeni hâline geldiğini göstermektedir.

Dil Yeterlilikleri ve Öğrenen Özerkliği Üzerindeki Etkiler

Bulgular, ÜYZ araçlarının yalnızca öğrencilerin dil yeterlilikleri değil, aynı zamanda duyuşsal ve bilişsel yönlerini de anlamlı biçimde etkilediğini ortaya koymaktadır. Bu araçlarla kurulan etkileşim, özellikle konuşma kaygısını azaltmakta, öğrencilerin daha bağımsız, öz düzenlemeye dayalı öğrenme stratejilerinin gelişimini desteklemektedir. Öğrenme sürecinde bireyin sadece dilsel çıktıları değil, öğrenmeye dair tutumları da dönüştüren bu etki, ÜYZ araçlarının geleneksel öğretim yöntemlerine kıyasla daha bütüncül katkılar sunduğunu göstermektedir.

Sayed vd. (2024) tarafından yürütülen deneysel çalışmada, ChatGPT destekli konuşma pratiği yapılan bir grup öğrencinin sadece sözel üretim becerilerinde değil, aynı zamanda psikolojik iyi oluş düzeylerinde ve öğrenen özerkliklerinde anlamlı artışlar gözlemlenmiştir. Araştırmada kullanılan üç zamanlı test (ön test, son test, gecikmeli test) tasarımı, öğrencilerin sadece kısa vadede değil, uzun vadeli kazanımlar elde ettiğini ortaya koymuştur. Bu sonuçlar, ÜYZ araçlarının sürdürülebilir öğrenme çıktılarına ulaşmada da etkili olabileceğine işaret etmektedir. Benzer şekilde, Pangestu ve Suwartono (2024) tarafından gerçekleştirilen nitel çalışmada, dört üniversite öğrencisinin üretken yapay zekâ ile olan deneyimleri incelenmiştir. Katılımcılar, özellikle kelime bilgisi, telaffuz ve dilbilgisi alanlarında gelişim sağladıklarını ve bu gelişimlerin öğrenme motivasyonlarını artırdığını belirtmişlerdir. Ayrıca öğrenciler, bu araçlar sayesinde daha fazla pratik yapabildiklerini, hata yapmaktan çekinmediklerini ve böylece daha aktif öğrenen rollerine geçtiklerini ifade etmişlerdir.

Ancak her iki çalışmada da bazı sınırlılık alanları dikkat çekmiştir. Öğrenciler, YZ sistemlerinin bağlamı tam olarak anlayamaması, mizah ya da kültürel incelikleri algılayamaması gibi nedenlerle zaman zaman hayal kırıklığı yaşadıklarını belirtmişlerdir. Bu durum, yapay zekânın duyuşsal etkileşim kurma kapasitesinde hâlâ gelişime açık yönler bulunduğunu göstermektedir. Genel olarak değerlendirildiğinde, ÜYZ tabanlı uygulamaların öğrenenlerde özerklik ve iletişimsel yeterlik gelişimi açısından önemli fırsatlar sunduğu söylenebilir.

Konuşma Değerlendirmesinde Yenilikçi Teknik Yaklaşımlar

ÜYZ destekli sistemlerin gelişimi, konuşma becerilerinin değerlendirilmesinde yenilikçi yaklaşımların önünü açmıştır. Geleneksel ölçme değerlendirme uygulamalarında, genellikle konuşma akıcılığı, telaffuz doğruluğu ve kelime dağarcığı gibi temel parametreler öne çıkarken; günümüzde ÜYZ tabanlı sistemler, çok daha kapsamlı ve ayrıntılı analizleri mümkün kılmaktadır. Bu araçlar sayesinde yalnızca dilsel çıktılar değil, aynı zamanda öğrencilerin söylem organizasyonu, dil bilgisel yeterliği ve bağlamsal uygunluğu gibi üst düzey dil yeterlikleri de değerlendirilebilmektedir.

Kopparapu vd. (2024) tarafından geliştirilen sistem, konuşulan dilin dil bilgisel doğruluğunu analiz edebilen özgün bir değerlendirme modeli sunmaktadır. Bu sistemde ChatGPT tarafından oluşturulan özel paragraflar, öğrencilere sesli olarak okutturulmuş ve özelleştirilmiş dil modelleri kullanılarak dilbilgisi hataları otomatik şekilde tespit edilmektedir. Sistemin kullandığı özel ASR-CLM (Automatic Speech Recognition with Custom Language Model) yapısı, mevcut ses tanıma motorlarına kıyasla çok daha yüksek doğruluk oranlarına ulaşmış, bu da özellikle sözlü dilde dil bilgisel yeterliğin değerlendirilmesinde önemli bir adım olarak değerlendirilmiştir. Sistem, ChatGPT'nin her aday için güçlük düzeyi denk fakat benzersiz paragraflar üretmesi sayesinde sınav maddelerinin ezberlenerek öğretilabilir hâle gelmesini önlemiş; madde tekrarını/sızıntısını azaltarak ölçmenin güvenilirliğini ve ayırt ediciliğini güçlendirmiştir. Buna paralel olarak, Ma vd. (2025) çalışmalarında multimodal büyük dil modelleriyle yürütülen değerlendirme süreçlerinde hem sınıflama hem de regresyon tabanlı tekniklerin başarıyla kullanılabildiği gösterilmiştir. Araştırma, sınırlı veriyle dahi etkili sonuçlar alınabildiğini ve bu modellerin çapraz görevlerde yüksek genelleme performansı sergileyebildiğini ortaya koymuştur. Özellikle eğitimci-öğrenci arasındaki puanlama farklarını azaltma ve değerlendirme sürecinde ortaya çıkabilecek öznellik sorunlarını ortadan kaldırma potansiyeli ön plana çıkmaktadır.

Tüm bu bulgular, ÜYZ destekli modellerin, konuşma değerlendirme süreçlerinde sadece zaman ve emek tasarrufu sağlamadığını; aynı zamanda daha bütüncül, nesnel ve ölçeklenebilir değerlendirme sistemlerinin kurulmasını mümkün kılabileceğine bir işaret olarak kabul edilebilecek bulgulardır.

Sistemsel Sınırlılıklar ve Etik Meseleler

Her ne kadar ÜYZ sistemleri dil öğrenme ve değerlendirme süreçlerinde önemli avantajlar sunsa da, mevcut çalışmalar bu teknolojilerin pedagojik faydalarının yanı sıra çeşitli sınırlılık ve etik riskleri de beraberinde getirdiğini göstermektedir. Özellikle öğretmen ve öğrencilerin deneyimlerine dayalı niteliksel veriler, yapay zekâ destekli konuşma değerlendirme sistemlerinin hâlen çözülmesi gereken çok sayıda yapısal ve uygulamaya dönük sorun barındırdığını ortaya koymaktadır.

Muthmainnah vd. (2025) tarafından yürütülen çalışma, ChatGPT'nin sunduğu erişilebilirlik, esneklik ve bireyselleştirme olanaklarının öğrenme süreçlerini desteklediğini kabul etmekle birlikte; bu sistemlerin insan benzeri bilişsel esneklikten, kültürel farkındalık ve duygusal zekâdan yoksun olduğunu vurgulamaktadır. Araştırma, özellikle duygusal geri bildirim gerektiren iletişim senaryolarında yapay zekânın yetersiz kaldığını ve bunun öğrencilerle kurulan etkileşimin doğallığını zedeleyebileceği sonucuna ulaşmıştır. Bu noktada, yapay zekânın bir nevi otomatik öğretmen gibi konumlandırılmaması gerektiği; bunun yerine eğitimi tamamlayıcı bir destek unsuru olarak değerlendirilmesinin daha işlevsel olacağı ifade edilmiştir. Benzer şekilde, Pangestu ve Suwartono (2024) tarafından gerçekleştirilen çalışmada da yapay zekâ sistemlerinin özellikle etkileşimde bağlamsal sınırlamalara sahip olduğu; mizahi ifadeleri anlamakta güçlük çektiğini ve karmaşık durumları kavrayamadığı belirtilmiştir. Katılımcılar, bazı durumlarda yapay zekâ ile etkileşimlerinde beklentileri karşılamayan yanıtlar, katı ve sınırlı etkileşim imkânları ve özellik kısıtlamaları nedeniyle zorluk yaşadıklarını ifade etmişlerdir. Bu durumların, öğrenme sürecinin verimini azalttığına dikkat çekilmiştir.

Etik açıdan da önemli soru işaretleri gündeme gelmektedir. Özellikle öğrencilerin kişisel verilerinin gizliliği, veri güvenliği, akademik dürüstlük ilkelerine uyum gibi konular (Ma, 2024; Muthmainnah vd., 2025) yapay zekâ tabanlı değerlendirme sistemlerinin geniş ölçekte uygulanabilirliğini sınırlayan faktörler arasındadır.

Bu bağlamda, mevcut sistemlerin daha kullanıcı dostu, eğitsel olarak bütünleştirilmiş ve etik kurallar çerçevesinde yapılandırılmış sürümlere ihtiyaç olduğu açıktır. ÜYZ'nin potansiyelinden maksimum düzeyde yararlanabilmek için, bu teknolojilerin eğitsel hedeflerle uyumlu hâle getirilmesi, kültürel ve duygusal bağlamları dikkate alan tasarım ilkeleriyle yeniden ele alınması önem taşımaktadır.

Tablo 3. Tematik Bulgular Tablosu

Tema	Öne Çıkan Bulgular	Destekleyen Çalışmalar
Otomatik Geribildirim ve Bireyselleştirilmiş Değerlendirme	Üretken yapay zekâ araçları, konuşma becerisi değerlendirmesinde anlık, ayrıntılı ve kişiselleştirilmiş geribildirim sunarak bireysel öğrenme süreçlerini desteklemektedir. Bu araçlar öğretmen yükünü azaltmakta, değerlendirme sürecini nesnelleştirmekte ve çok boyutlu performans analizi sağlamaktadır.	Gaballo (2024); Ma (2024); Sayed vd. (2024)
Dil Yeterlilikleri ve Öğrenen Özerkliği Üzerindeki Etkiler	ÜYZ araçları, sadece dil yeterliğini değil, öğrenenlerin konuşma kaygısı ve öğrenen özerkliği gibi duygusal faktörlerini de olumlu yönde etkilemektedir. Öğrencilerin psikolojik iyi oluşları desteklenmekte ve öğrenme motivasyonları artmaktadır.	Sayed vd. (2024); Pangestu ve Suwartono (2024)
Konuşma Değerlendirmesinde Yenilikçi Teknik Yaklaşımlar	ÜYZ sistemleri, konuşma girdilerinin otomatik değerlendirilmesinde sınıflama, regresyon ve dil bilgisel hata tespiti gibi gelişmiş yöntemler sunmakta; bu durum nesnellik, tutarlılık ve ölçeklenebilirlik açısından geleneksel yöntemlere alternatif oluşturmaktadır.	Kopparapu vd. (2024); Ma vd. (2025)
Sistemsel Sınırlılıklar ve Etik Meseleler	ÜYZ sistemlerinin bağlamsal farkındalık, insani sezgi ve esneklik gerektiren durumlarda sınırlı olduğu; teknik, etik ve eğitsel sorunların dikkatle ele alınması gerektiği vurgulanmaktadır. Aşırı bağımlılık, yanlış yönlendirme ve kültürel duyarsızlık potansiyel riskler arasındadır.	Ma (2024), Muthmainnah vd. (2025); Pangestu ve Suwartono (2024)

Tartışma ve Sonuç

Çalışmanın bulguları, ÜYZ destekli araçların yabancı dil öğrenen bireylerin konuşma becerilerinin değerlendirilmesinde çok boyutlu işlevlere sahip olduğunu ortaya koymaktadır. Tematik analiz neticesinde elde edilen veriler, değerlendirme süreçlerinin yalnızca teknik doğrulukla sınırlı kalmadığını; öğrenen deneyimi, psikolojik iyi oluş, pedagojik fayda ve etik/sistemsel boyutları da kapsayacak şekilde geniş bir çerçevede ele alınması gerektiğini göstermektedir. Bu bağlamda, ulaşılan bulgular dört ana tema altında toplanmıştır: otomatik geribildirim ve bireyselleştirilmiş değerlendirme, dil yeterlilikleri ve öğrenen özerkliği üzerindeki etkiler, teknik yenilikler ve sistemsel sınırlılıklar.

ÜYZ tabanlı sistemlerin bu konudaki en dikkat çekici katkılarından biri, öğrencilere bireysel geribildirim sunma kapasiteleridir. Büyük dil modellerinin sağladığı anlık, ayrıntılı ve hedefe yönelik geribildirimler, öğrencilerin bireysel gelişim sürecine önemli katkılar sunmaktadır. Gaballo (2024), geniş ölçekli bir lisans sınıfında ÜYZ araçlarının öğretim üyelerinin değerlendirme yükünü önemli ölçüde azalttığını ve değerlendirme süreçlerini daha sistematik, tutarlı ve adil bir hâle getirdiğini ortaya koymuştur. Benzer biçimde, Ma (2024) çalışmasında ChatGPT'nin kişiselleştirilmiş konuşma pratiği ve otomatik geribildirim işlevleri sayesinde orta düzeyde Çince öğrenen öğrencilerin sözlü yeterliklerinde anlamlı gelişmeler sağlandığı görülmüştür.

Sayed vd. (2024) ise ChatGPT'nin TOEFL iBT konuşma bölümünü değerlendirme sürecine entegre edildiği deneysel çalışmasında, bireysel geribildirimlerin hem kısa vadeli hem de uzun vadeli dil gelişimi üzerinde etkili olduğunu göstermektedir. Bu bulgular, ÜYZ sistemlerinin sadece teknolojik yenilik olarak değil, aynı zamanda öğretimsel değeri yüksek eğitsel araçlar olarak değerlendirilebileceğini göstermektedir.

ÜYZ sistemlerinin konuşma becerisi üzerindeki etkisi yalnızca teknik değerlendirme ile sınırlı kalmamakta; aynı zamanda öğrencilerin öğrenmeye yönelik duyuşsal ve psikolojik durumlarını da olumlu yönde etkilemektedir. Sayed vd. (2024), ChatGPT destekli konuşma pratiği uygulayan öğrencilerde yalnızca akademik performans değil; psikolojik iyi oluş ve özerklik düzeylerinde de anlamlı artışlar olduğunu bildirmiştir. Bu sonuç, ÜYZ araçlarının öğrenenlerde daha bağımsız, öz düzenleyici ve güvenli bir öğrenme ortamı oluşturduğunu göstermektedir. Benzer şekilde, Pangestu ve Suwartono (2024) çalışması da öğrencilerin kişisel deneyimlerini merkeze alarak, ÜYZ araçlarının özellikle telaffuz, sözcük bilgisi ve dil bilgisel yapıların ediniminde etkili olduğunu; ayrıca öğrenen motivasyonunu artırdığını belirtmiştir. Bununla birlikte, YZ'nin insan etkileşimini ikame edemeyeceğini, bağlamsal anlam ve mizah gibi nüansları yeterince yakalayamadığını ifade etmişlerdir. Bu bağlamda, ÜYZ sistemlerinin dil yeterliği gelişimi kadar, öğrenme sürecindeki özerklik düzeylerini destekleyen bir yapıya sahip olduğu görülmektedir.

ÜYZ tabanlı sistemler, geleneksel ölçme-değerlendirme yöntemlerinin ötesine geçerek daha gelişkin, veri temelli ve çok katmanlı değerlendirme biçimleri sunmaktadır. Kopparapu vd. (2024), geleneksel SLA sistemlerinin aksine dil bilgisel doğruluğu da ölçen bir model geliştirerek, öğrenci konuşmalarını özgün olarak oluşturulan metinlerle karşılaştırmış ve özelleştirilmiş dil modelleri sayesinde yüksek doğrulukla hataları saptayabilmiştir. Bu model, sözlü üretim sırasında dil bilgisel doğruluğun da nesnel şekilde ölçülebileceğini ortaya koymaktadır. Ayrıca Ma vd. (2025) çalışması, multimodal büyük dil modelleri kullanarak konuşma değerlendirmesini sınıflama ve regresyon temelli skorlarla gerçekleştirmiştir. Bu sistemlerin hem kontrollü hem de serbest konuşma türlerinde yüksek doğruluk oranlarına ulaştığı, sınırlı veriyle dahi genelleme yeteneklerinin güçlü olduğu belirtilmiştir. Bu durum, konuşma değerlendirmesindeki önyargı, geçerlilik ve tutarlılık gibi sorunların üstesinden gelme potansiyeline işaret etmektedir.

Her ne kadar ÜYZ sistemleri teknik ve eğitsel anlamda güçlü katkılar sunsa da, bu sistemlerin belirli bilişsel, etik ve işlevsel sınırlılıkları göz ardı edilmemelidir. Muthmainnah vd. (2025), çalışmalarında ChatGPT'nin öğrencilerin dil performansını desteklemede etkili olduğunu belirtirken; sistemin insan benzeri bilişsel yeterlikten yoksun oluşunun, bağlamsal duyarlılık ve duygusal zekâ eksikliğinin önemli sınırlılıklar yarattığını vurgulamışlardır. Araştırmacılar, bu sistemlerin öğretmenleri ikame etmek yerine, tamamlayıcı bir kaynak olarak konumlandırılması gerektiği yönünde görüş bildirmiştir. Benzer eleştiriler, Pangestu ve Suwartono (2024) çalışmasında da görülmektedir. Katılımcılar, YZ araçlarının bazen anlamsal belirsizlikler yaratabildiğini, karmaşık ya da kültürel bağlamlara yeterince duyarlı olmadığını ve teknik yetersizliklerin öğrenme deneyimini olumsuz etkileyebileceğini ifade etmişlerdir. Bu durum, ÜYZ araçlarının daha etkili şekilde entegre edilebilmesi için hem teknolojik hem de eğitsel düzeyde yapılandırılması gerektiğini göstermektedir.

Öneriler

Bu derlemede analiz edilen çalışmalar, ÜYZ destekli araçların yabancı dil öğrenen bireylerin konuşma becerilerinin değerlendirilmesinde ve bu becerilerin üzerindeki etkilerine dair önemli bulgular sunmakla birlikte, alanın henüz gelişiminin erken evresinde olduğunu da açıkça göstermektedir. Bu bağlamda, gelecekte gerçekleştirilecek araştırmalarda aşağıdaki öneriler dikkate alınabilir.

Sayed vd. (2024) çalışmasında, ChatGPT destekli konuşma pratiğinin öğrenci başarısı ve psikolojik iyi oluş üzerindeki etkileri boylamsal bir tasarımla ele alınmıştır. Ancak bu tür uzun süreli izlemelerin sınırlı sayıda katılımcıyla gerçekleştirilmesi, genelleme yapılabilirliği kısıtlamaktadır. Gelecek çalışmalarda, daha geniş örneklemle dayanan çok-merkezli araştırmalarla farklı yaş grupları, dil düzeyleri ve sosyo-kültürel bağlamlar arasında karşılaştırmalar yapılması önerilmektedir.

ÜYZ araçlarının yalnızca bilişsel çıktılar değil, duyuşsal değişkenler üzerinde de etkili olduğu belirlenmiştir (Sayed vd., 2024; Pangestu & Suwartono, 2024). Ancak psikolojik iyi oluş, konuşma kaygısı, motivasyon ve öğrenen özerkliği gibi değişkenlere dair bulgular büyük ölçüde nitel verilere dayanmaktadır. Bu nedenle, bu değişkenlerin nicel ölçeklerle izlenmesi önem taşımaktadır.

Ma vd. (2025) ve Kopparapu vd. (2024) tarafından geliştirilen büyük dil modeli tabanlı otomatik değerlendirme sistemleri, geleneksel değerlendirme yöntemleriyle karşılaştırıldığında güçlü bir potansiyel sunmaktadır. Ancak bu sistemlerin güvenilirliği, doğruluğu ve öğretici puanlamalarla tutarlılığı henüz sınırlı ölçekte değerlendirilmiştir. Gelecek çalışmalar, insan puanlayıcılarla karşılaştırmalı kodlama çalışmaları ve uyum katsayıları ile daha sistematik biçimde yapılması bu sistemlerin eğitim bağlamlarında etkin biçimde kullanılabilirliğini değerlendirmek açısından önemli gözükmektedir.

Muthmainnah vd. (2025) ve Pangestu & Suwartono (2024), ÜYZ sistemlerinin kültürel bağlamlara yeterince duyarlı olmadığını ve mizah, ima gibi dilin bağlama bağlı yönlerini kavramada yetersiz kaldığını vurgulamaktadır. Bu durum, özellikle iletişimsel beceriler açısından önemli bir sınırlılık oluşturmaktadır. İleri araştırmalar, ÜYZ tabanlı sistemlerin kültürel varyasyonlara ve çok dilli ortamlarda kullanılabilirliğine ilişkin dil çözümlemeleriyle desteklenebilir.

Çalışmalarda YZ araçlarının bilişsel sınırlarının yanı sıra, akademik etik, yapay içerik üretimi ve akademik dürüstlük konularında belirsizlikler oluşturduğu ifade edilmiştir. Bu bağlamda, gelecekteki araştırmalar sadece eğitsel işlevselliği değil, aynı zamanda etik ve yapay zekâ okuryazarlığı boyutlarını da odağa alarak daha bütüncül bir çerçeve geliştirebilir.

Kopparapu vd. (2024) tarafından önerilen özel amaçlı dil modelleri ve ASR sistemlerinin başarısı, mühendislik ile dil pedagojisi arasında disiplinler arası bir iş birliği gerekliliğini ön plana çıkarmaktadır. Gelecek araştırmalarda, yapay zekâ mühendisliği, dil öğretimi, eğitim teknolojileri ve etik çalışma alanlarının ortaklaşa yürüteceği, kullanıcı merkezli tasarım süreçlerine dayanan deneysel çalışmalar önceliklendirilmelidir.

Bu hızlı derleme, dâhil edilme ölçütlerini karşılayan yalnızca yedi ampirik çalışmaya dayanmakta ve kanıtlar 2024-2025 dönemine ait olduğundan bulgular erken evre, dar bir kanıt gövdesini temsil etmektedir. Örneklemeler ağırlıklı yükseköğretim bağlamlarında yoğunlaşmış, K-12 (okul öncesi-lise), farklı diller ve sosyo-kültürel bağlamlar yetersiz temsil edilmiştir; ayrıca birçok çalışmada kısa vadeli sonuçlar raporlanmış, uzunlamasına kanıt sınırlı kalmıştır.

Çıkar Çatışması Beyanı

Yazar, bu makalenin araştırma, yazarlık ve/veya yayın süreci ile ilgili herhangi bir potansiyel çıkar çatışması olmadığını beyan eder.

Mali Destek

Yazar bu makalenin araştırılması, yazılması ve/veya yayınlanması için herhangi bir mali destek almamıştır.

Yayın Etiği Beyanı

Çalışmada etik dışı bir husus bulunmadığını, araştırma ve yayın etiğine özenle uyulduğunu beyan ederim.

Yazar Katkı Oranı

Çalışma, tek yazar tarafından yürütülmüş ve raporlanmıştır.

Etik Kurul İzni

Bu çalışma, yalnızca daha önce yayımlanmış verilerin analizine dayanan bir hızlı derleme niteliğindedir. Herhangi bir bireysel veri toplanmadığı, insan veya hayvan deneklerle doğrudan bir etkileşim söz konusu olmadığı için etik kurul onayı gerektirmemektedir.

Kaynakça

- Ayeni, O. O., Al Hamad, N. M., Chisom, O. N., Osawaru, B., & Adewusi, O. E. (2024). AI in education: A review of personalized learning and educational technology. *GSC Advanced Research and Reviews*, 18(2), 261-271. <https://doi.org/10.30574/gscarr.2024.18.2.0062>
- Bae, M., Wang, J., Xue, H., Chong, S. M., Kwon, O., & Ki, C. W. (2024). Does ChatGPT help or hinder education? Exploring its benefits, challenges, student guilt, and the need for educator training. *International Journal of Fashion Design, Technology and Education*, 1-16. <https://doi.org/10.1080/17543266.2024.2430585>
- Bolton, C., Machová, V., Kovacova, M., & Valaskova, K. (2018). The power of human-machine collaboration: Artificial intelligence, business automation, and the smart economy. *Economics, Management, and Financial Markets*, 13(4), 51-56.
- Celik, B., Yildiz, Y., & Kara, S. (2025). Using ChatGPT as a virtual speaking tutor to boost EFL learners' speaking self-efficacy. *Australian Journal of Applied Linguistics*, 8(1). <https://doi.org/10.29140/ajal.v8n1.102418>
- Dai, M. (2022). Intelligent Correction System of Students' English Pronunciation Errors Based on Speech Recognition Technology. *Journal of Information & Knowledge Management*, 21(Supp02), 2240013. <https://doi.org/10.1142/S0219649222400135>
- Derner, E., & Batistič, K. (2023). Beyond the safeguards: Exploring the security risks of chatgpt. *arXiv preprint* <https://doi.org/10.48550/arXiv.2305.08005>
- Dobbins, M. (2017). Rapid review guidebook. *Natl Collab Cent Method Tools*, 13, 25.
- Dündar, A. E. (2024). *Création et évaluation de l'efficacité d'un système tutoriel intelligent pour l'enseignement de la grammaire du FLE*. (Yayın No. 914033) [Doktora tezi, Anadolu Üniversitesi]. YÖK Ulusal Tez Merkezi.
- Dündar, O. İ. (2024). Utilisation Potentielle De Chatgpt Dans L'apprentissage Des Langues Etrangères: Exploration Des Possibilités Selon Les Niveaux Langagiers Du CECRL. *Kabramanmaraş Sütçü İmam Üniversitesi Sosyal Bilimler Dergisi*, 21(1), 63-75. <https://doi.org/10.33437/ksusbd.1384040>
- Dündar, O. İ., & Aydoğu, C. (2024). Création d'un Agent Conversationnel Pour Une Classe De Français Langue Etrangère: Les Effets Sur L'autonomie De L'apprenant. *Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 33(2), 750-762. <https://doi.org/10.35379/cusosbil.1411764>
- Dündar, O. İ., & Aydoğu, C. (2024). Rencontrer l'Autre de la perspective actionnelle via l'intelligence artificielle. *Balıkesir Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 27(51), 1-16. <https://doi.org/10.31795/baunsobed.1415369>
- Gaballo, V. (2024, November). Revolutionizing language teaching: AI in oral language assessment. In *Conference Proceedings. Innovation in Language Learning 2024*.
- Gammoh, L. A. (2024). ChatGPT in academia: Exploring university students' risks, misuses, and challenges in Jordan. *Journal of Further and Higher Education*, 48(6), 608-624. <https://doi.org/10.1080/0309877X.2024.2378298>
- Görgüt, R. Ç. (2024). Yapay zekâ ve matematik eğitimi. *Eğitim ve Bilim 2023-IV*, 43.
- Halagatti, M., Gadag, S., Mahantshetti, S., Hiremath, C. V., Tharkude, D., & Banakar, V. (2023). Artificial intelligence: the new tool of disruption in educational performance assessment. In *Smart analytics*,

- artificial intelligence and sustainable performance management in a global digitalised economy* (pp. 261-287). Emerald Publishing Limited. <https://doi.org/10.1108/S1569-37592023000110A014>
- Hashim, S., Omar, M. K., Ab Jalil, H., & Sharef, N. M. (2022). Trends on technologies and artificial intelligence in education for personalized learning: systematic literature. *Journal of Academic Research in Progressive Education and Development*, 12(1), 884-903. <http://dx.doi.org/10.6007/IJARPED/v11-i1/12230>
- Hawanti, S., & Zubayduloevna, K. M. (2023). AI chatbot-based learning: alleviating students' anxiety in English writing classroom. *Bulletin of Social Informatics Theory and Application*, 7(2), 182-192. <https://doi.org/10.31763/businta.v7i2.659>
- Hooda, M., Rana, C., Dahiya, O., Rizwan, A., & Hossain, M. S. (2022). Artificial intelligence for assessment and feedback to enhance student success in higher education. *Mathematical Problems in Engineering*, 2022(1). <https://doi.org/10.1155/2022/5215722>
- İnal, E. (2020). Yabancı dil öğretiminde konuşma becerisinin geliştirilmesi ve ölçme değerlendirme süreçlerine yönelik düşünceler: öğretici yeterlikleri. *Aydın TÖMER Dil Dergisi*, 5(2), 189-204. https://doi.org/10.17932/IAU.TOMER.2016.019/tomer_v05i2003
- Jeon, J., Lee, S., & Choe, H. (2023). Beyond ChatGPT: A conceptual framework and systematic review of speech-recognition chatbots for language learning. *Computers & Education*, 206. <https://doi.org/10.1016/j.compedu.2023.104898>
- Khatiri, B. B., & Karki, P. D. (2023). Artificial intelligence (AI) in higher education: Growing academic integrity and ethical concerns. *Nepalese Journal of Development and Rural Studies*, 20(01), 1-7. <https://doi.org/10.3126/njdrs.v20i01.64134>
- Kopparapu, S. K., Bhat, C., & Panda, A. (2024). Spoken grammar assessment using LLM. *arXiv preprint* <https://doi.org/10.48550/arXiv.2410.01579>
- Kumar, S., Musharaf, D., Musharaf, S., & Sagar, A. K. (2023, June). A comprehensive review of the latest advancements in large generative AI models. In *International Conference on Advanced Communication and Intelligent Systems* (pp. 90-103). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-45121-8_9
- Lu, Y. (2019). Artificial intelligence: a survey on evolution, models, applications and future trends. *Journal of Management Analytics*, 6(1), 1-29. <https://doi.org/10.1080/23270012.2019.1570365>
- Ma, L. (2024). Research on Oral Teaching of International Chinese Language Education Assisted by ChatGPT. *Journal of Higher Vocational Education (ISSN: 3005-5784)*, 1(2), 129.
- Ma, R., Qian, M., Tang, S., Bannò, S., Knill, K. M., & Gales, M. J. (2025). Assessment of L2 Oral Proficiency using Speech Large Language Models. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2505.21148>
- Mahmoud, C. F., & Sørensen, J. T. (2024). Artificial Intelligence in Personalized Learning with a Focus on Current Developments and Future Prospects. *Research and Advances in Education*, 3(8), 25-31. <https://doi.org/10.56397/RAE.2024.08.04>
- Malfatti, F. I. (2025). Chatgpt, education, and understanding. *Social Epistemology*, 1-15. <https://doi.org/10.1080/02691728.2025.2449599>
- Mema, B., Basholli, F., & Hyka, D. (2024). Learning transformation and virtual interaction through ChatGPT in Albanian higher education. *Advanced Engineering Science*, 4, 130-140.
- Muthmainnah, M., Asad, M. M., Yakin, A. A., & Almusharraf, N. M. (2025). Human-robot interaction using ChatGPT for technology driven language learning: contextual insights from the higher education institution of Indonesia. *Asian Education and Development Studies*. <https://doi.org/10.1108/AEDS-01-2025-0038>
- Murtiningsih, S., Sujito, A., & Soe, K. K. (2024). Challenges of using ChatGPT in education: A digital pedagogy analysis. *Int J Eval & Res Educ ISSN*, 2252(8822), 3467. <https://doi.org/10.11591/ijere.v13i5.29467>
- Najafov, E. (2024). Overview of ChatGPT. In *ChatGPT for Marketing: Learn Practical Applications of ChatGPT for Marketing* (pp. 1-15). Berkeley, CA: Apress. https://doi.org/10.1007/979-8-8688-0312-3_1

- Neupane, K. R. (2025). *Strategies of Teaching Speaking Skills in Secondary English Classroom: An Auto/Ethnographic Study* [Yüksek lisans tezi, Katmandu Üniversitesi]. Katmandu University Digital Repository.
- Owan, V. J., Abang, K. B., Idika, D. O., Etta, E. O., & Bassey, B. A. (2023). Exploring the potential of artificial intelligence tools in educational measurement and assessment. *Eurasia journal of mathematics, science and technology education*, 19(8). <https://doi.org/10.29333/ejmste/13428>
- Page, L. C., & Gehlbach, H. (2017). How an artificially intelligent virtual assistant helps students navigate the road to college. *Aera Open*. <https://doi.org/10.1177/2332858417749220>
- Pangestu, H. P., & Suwartono, T. (2024). Exploring the Role of Artificial Intelligence (AI) Support in Assisting Students' English-Speaking Skills. *Educalitra: English Education, Linguistics, and Literature Journal*, 3(2), 134-147. <https://doi.org/10.5281/zenodo.13134979>
- Park, D. M., & Lee, H. J. (2024). Literature Review of AI Hallucination Research Since the Advent of ChatGPT: Focusing on Papers from arXiv. *Informatization Policy*, 31(2), 3-38. <https://doi.org/10.22693/NIAIP.2024.31.2.003>
- Salas-Pilco, S. Z., Xiao, K., & Hu, X. (2022). Artificial intelligence and learning analytics in teacher education: A systematic review. *Education Sciences*, 12(8), 569. <https://doi.org/10.3390/educsci12080569>
- Sayed, B. T., Bani Younes, Z. B., Alkhayyat, A., Adhamova, I., & Teferi, H. (2024). To be with artificial intelligence in oral test or not to be: a probe into the traces of success in speaking skill, psychological well-being, autonomy, and academic buoyancy. *Language Testing in Asia*, 14(1), 49. <https://doi.org/10.1186/s40468-024-00321-0>
- Schwalbe, N., & Wahl, B. (2020). Artificial intelligence and the future of global health. *The Lancet*, 395(10236), 1579-1586.
- Sreen, A. S., & Majid, M. H. M. (2024). Leveraging ChatGPT for personalized learning: A systematic review in educational settings. *Amandemen: Journal of Learning, Teaching and Educational Studies*, 2(1), 63-78. <https://doi.org/10.61166/amd.v2i1.72>
- Stahl, B. C., & Eke, D. (2024). The ethics of ChatGPT—Exploring the ethical issues of an emerging technology. *International Journal of Information Management*, 74, 102700. <https://doi.org/10.1016/j.ijinfomgt.2023.102700>
- Sullivan, M., Kelly, A., & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning & Teaching*, 6(1), 1-10. <https://doi.org/10.37074/jalt.2023.6.1.17>
- Surden, H. (2018). Artificial intelligence and law: An overview. *Ga. St. UL Rev.*, 35, 1305.
- Tapalova, O., & Zhiyenbayeva, N. (2022). Artificial intelligence in education: AIED for personalised learning pathways. *Electronic Journal of e-Learning*, 20(5), 639-653.
- Terzopoulos, G., & Satratzemi, M. (2019, September). Voice assistants and artificial intelligence in education. In *Proceedings of the 9th Balkan Conference on Informatics* (pp. 1-6). <https://doi.org/10.1145/3351556.3351588>
- Vetrivel, S. C., Arun, V. P., Ambikapathi, R., & Saravanan, T. P. (2025). Automated Grading Systems: Enhancing Efficiency and Consistency in Student Assessments. In *Adopting Artificial Intelligence Tools in Higher Education* (pp. 41-61). CRC Press.
- Wang, Y. (2025). A Study on the Efficacy of ChatGPT-4 in Enhancing Students' English Communication Skills. *SAGE Open*, 15(1). <https://doi.org/10.1177/21582440241310644>
- Xu, S., Chen, P., & Zhang, G. (2024). Exploring the impact of the use of ChatGPT on foreign language self-efficacy among Chinese students studying abroad: The mediating role of foreign language enjoyment. *Heliyon*, 10(21). <https://doi.org/10.1016/j.heliyon.2024.e39845>
- Xu, W., Meng, J., Raja, S. K. S., Priya, M. P., & Kiruthiga Devi, M. (2023). Artificial intelligence in constructing personalized and accurate feedback systems for students. *International Journal of Modeling, Simulation, and Scientific Computing*, 14(01), 2341001. <https://doi.org/10.1142/S1793962323410015>

- Xu, X., Wang, X., Zhang, Y., & Zheng, R. (2024). Applying ChatGPT to tackle the side effects of personal learning environments from learner and learning perspective: An interview of experts in higher education. *PLoS One*, 19(1). <https://doi.org/10.1371/journal.pone.0295646>
- Yenduri, G., Srivastava, G., Maddikunta, P. K. R., Jhaveri, R. H., Wang, W., Vasilakos, A. V., & Gadekallu, T. R. (2023). Generative pre-trained transformer: A comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions. <https://doi.org/10.48550/arXiv.2305.10435>
- Zaghlool, Z. D., & Khasawneh, M. A. S. (2023). Incorporating the impacts and limitations of AI-driven feedback, evaluation, and real-time conversation tools in foreign language learning. *Migration Letters*, 20(7).
- Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review. *Smart Learning Environments*, 11(1), 28. <https://doi.org/10.1186/s40561-024-00316-7>
- Zou, B., Liviero, S., Ma, Q., Zhang, W., Du, Y., & Xing, P. (2024). Exploring EFL learners' perceived promise and limitations of using an artificial intelligence speech evaluation system for speaking practice. *System*, 126. <https://doi.org/10.1016/j.system.2024.103497>

EXTENDED SUMMARY

This extended summary provides an overview of a rapid review study that investigates the role of generative artificial intelligence (GAI) tools in assessing the speaking proficiency of foreign language learners. As AI technologies, especially generative models like ChatGPT, become increasingly embedded in educational contexts, understanding their pedagogical value and limitations has become essential. The study uses a thematic analysis method to synthesize recent empirical findings, offering insights into how GAI tools are shaping the landscape of language assessment.

A comprehensive literature search was conducted across four major academic databases, Scopus, Google Scholar, Web of Science, and EBSCO, targeting publications between 2018 and 2025. Sixty-seven relevant studies were initially identified, and based on predefined inclusion criteria, seven empirical studies that examined the use of GAI tools in speaking skill assessment for foreign language learners were selected for in-depth analysis.

The thematic synthesis of the selected studies revealed four major themes:

Automated Feedback and Personalized Assessment: GAI tools have demonstrated a strong capability to provide instant, detailed, and tailored feedback to learners during oral production tasks. Large Language Models (LLMs) and Automatic Speech Recognition (ASR) systems, for instance, can detect pronunciation, fluency, grammatical errors, and vocabulary usage in real time. Such detailed feedback enables learners to identify and address their weaknesses independently. Gaballo (2024) showed how the integration of AI in a large undergraduate classroom significantly reduced the instructor's grading burden while enhancing the quality of learner feedback.

Linguistic Proficiency and Learner Autonomy: The reviewed studies suggest that GAI tools contribute positively not only to linguistic outcomes but also to affective variables such as motivation, speaking confidence, and learner autonomy. For example, Sayed et al. (2024) found that learners using ChatGPT in speaking practice reported higher levels of psychological well-being and self-directed learning. Similarly, Pangestu and Suwartono (2024) noted that students felt less anxious about making mistakes when interacting with AI, allowing for more frequent and fearless practice.

Innovative Technical Approaches: GAI tools enable novel assessment techniques that go beyond traditional human-based evaluation. Some studies introduced sophisticated scoring algorithms based on classification and regression models integrated with LLMs. For instance, Koppurapu et al. (2024) developed a customized ASR system to assess spoken grammar, providing a more nuanced understanding of learner performance. Ma et al. (2025) demonstrated the successful use of multimodal models that combine voice, text, and contextual data to assess speaking skills with higher accuracy and consistency.

Systemic Limitations and Ethical Concerns: Despite the promising advantages, GAI-based tools face notable challenges. One recurring issue is the lack of contextual and cultural awareness, which can result in inappropriate or misleading feedback. Furthermore, ethical concerns such as data privacy, algorithmic transparency, and the potential for overreliance on AI systems are also raised. Muthmainnah et al. (2025) emphasized that while AI offers scalability, it cannot replace the empathetic, culturally responsive support that human educators provide. Concerns were also raised about the capacity of GAI tools to handle humor, idiomatic expressions, or emotionally nuanced language.

Overall, the findings indicate that GAI tools, when used thoughtfully, can offer transformative benefits for speaking skill development in foreign language education. Their ability to provide real-time, customized feedback allows learners to monitor progress continuously and adjust their strategies accordingly. Moreover, these tools can alleviate some of the logistical and time-related burdens on educators, making speaking assessment more scalable and accessible.

However, for GAI integration to be truly effective, certain conditions must be met. First, educators should receive training in AI literacy to interpret AI-generated feedback and integrate it into their instructional design. Second, developers must work on enhancing the cultural and contextual sensitivity of GAI systems to ensure more authentic and meaningful interactions. Third, institutions must implement clear ethical guidelines concerning data use, academic integrity, and learner autonomy.

In conclusion, this rapid review highlights the dual nature of GAI in speaking assessment: it is both an opportunity and a challenge. As the technology continues to evolve, future research should prioritize longitudinal and mixed-method studies to explore long-term learning outcomes, cross-cultural applicability, and best practices for responsible implementation. Interdisciplinary collaboration among linguists, educators, computer scientists, and ethicists will be key to maximizing the pedagogical potential of generative AI while addressing its limitations.

By offering an evidence-based, thematically structured synthesis, this extended summary contributes to a deeper understanding of how generative AI can reshape the assessment of speaking proficiency in language learning contexts. It encourages a balanced, critical approach that embraces innovation while safeguarding pedagogical and ethical standards.