

ANKARA ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

EĞİTİM BİLİMLERİ ANABİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME PROGRAMI

AÇIK UÇLU MATEMATİK SORULARININ
DEĞERLENDİRİLMESİNDE YAPAY ZEKÂ VE
ÖĞRETMEN PUANLAMALARININ
KARŞILAŞTIRILMASI

YÜKSEK LİSANS TEZİ

SERRA SEZİN ŞENER

ANKARA
TEMMUZ, 2025



ANKARA ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

EĞİTİM BİLİMLERİ ANABİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME PROGRAMI

AÇIK UÇLU MATEMATİK SORULARININ
DEĞERLENDİRİLMESİNDE YAPAY ZEKÂ VE
ÖĞRETMEN PUANLAMALARININ
KARŞILAŞTIRILMASI

YÜKSEK LİSANS TEZİ

SERRA SEZİN ŞENER

PROF. DR. SEHER YALÇIN

ANKARA
TEMMUZ, 2025

Ankara Üniversitesi Eğitim Bilimleri Enstitüsü Müdürlüğüne,

Serra Sezin ŞENER adlı öğrencinin hazırladığı “AÇIK UÇLU MATEMATİK SORULARININ DEĞERLENDİRİLMESİNDE YAPAY ZEKÂ VE ÖĞRETMEN PUANLAMALARININ KARŞILAŞTIRILMASI” başlıklı bu çalışma Eğitim Bilimleri Anabilim Dalı / Eğitimde Ölçme ve Değerlendirme Programı’nda jüri üyelerince oy birliği ile **Yüksek Lisans Tezi** olarak kabul edilmiştir.

Jüri Üyeleri

İmza

Başkan

Doç. Dr. Münevver İlgün DİBEK

Üye

Prof. Dr. Seher YALÇIN

Üye

Prof. Dr. Celal Deha DOĞAN

ONAY

Bu tez Ankara Üniversitesi Lisansüstü Eğitim Öğretim Yönetmeliği’nin ilgili maddeleri uyarınca, jüri üyeleri tarafından .../.../20... tarihinde, Enstitü Yönetim Kurulu tarafından ise .../.../20... tarihinde kabul edilmiştir.

.....

Prof. Dr. Mehmet İkbâl YETİŞİR
Eğitim Bilimleri Enstitüsü Müdürü

ETİK İLKELERE UYGUNLUK BİLDİRİMİ

Tez içindeki bütün bilgileri akademik yazım kurallarına uygun biçimde raporlaştırdığımı ve bunları etik ilkelere (atıfta bulunulan tüm yapıtlara kaynaklarda yer verilmesi, tezde kullanılan bilgi ve belgelere resmi yollarla ulaşılması ve bunların aslı bozulmadan kullanılması vb.) uygun olarak elde ettiğimi ve sunduğumu bildiririm.

(İmza)

(Adı Soyadı)

X

ÖZET

AÇIK UÇLU MATEMATİK SORULARININ DEĞERLENDİRİLMESİNDE YAPAY ZEKÂ VE ÖĞRETMEN PUANLAMALARININ KARŞILAŞTIRILMASI

ŞENER, Serra Sezin

Yüksek Lisans Tezi, Eğitim Bilimleri Anabilim Dalı

Tez Danışmanı: Prof. Dr. Seher YALÇIN

Temmuz, 2025, xii + 101

Bu araştırma, yapay zeka (YZ) destekli sistemlerin matematik sorularının değerlendirilmesindeki yerini ve geleceğini öğretmen değerlendirmeleriyle karşılaştırmalı olarak incelemeyi amaçlamaktadır. Araştırmada karma yöntem benimsenmiş; bir özel okulun beş farklı yerleşkesinden (Ankara Oran, Ankara İncek, Antalya Muratpaşa, Antalya Manavgat, Sakarya) 217 7. sınıf öğrencisi, 4 ölçme uzmanı ve 5 matematik öğretmeni ile çalışılmıştır. Veri toplama sürecinde beş okulda aynı anda uygulanan 2. Dönem 1. ortak yazılısının 10 maddelik matematik testi kullanılmış, aynı öğrencilere ait yazılı kağıtları hem öğretmenler hem de YZ sistemi tarafından değerlendirilerek öğrencilere geri bildirimler verilmiştir. Nicel veriler güvenirlik, madde-toplam korelasyonu, ayırt edicilik ve ağırlıklandırılmış Kappa uyum analizleri ile incelenirken nitel veriler 36 öğretmenin katıldığı bir anket, yazılı sınavın uygulandığı okulun matematik öğretmenleri ve ölçme uzmanları ile yapılan birebir görüşmeler ve yazılı sonuçları doğrultusunda belirlenen farklı başarı seviyelerinden 30 öğrenci ile gerçekleştirilen odak grup görüşmeleri yoluyla toplanmıştır. Araştırmanın bulgularına göre, yazılı puanlamasında YZ sisteminin Cronbach alfa güvenirlik katsayısı öğretmen değerlendirmesinden yüksek çıkmıştır. Madde-toplam korelasyonları ve ayırt edicilik indekslerinde de YZ sisteminin daha tutarlı olduğu yönünde sonuçlar olmakla birlikte iki değerlendirmenin de güvenirliğinin yüksek olduğu belirlenmiştir. Ancak değerlendiriciler arası (öğretmen ve YZ) uyum analizinde algoritmik işlemler ve formül uygulamalarında yüksek (.60 ve üstünde) uyum gözlenirken, kavramsal anlama ve çok aşamalı problem çözme gerektiren sorularda düşük (.40 ve altında) uyum tespit edilmiştir. En büyük puan farklılıklarının yaşandığı maddeler; çok aşamalı yüzde hesaplamaları, gerçek hayat uygulamaları ve sözel ifadelerden denklem kurma gibi matematik konularını

içermektedir. Nitel bulgular, öğretmenlerin büyük çoğunluğunun YZ sistemlerinin eğitime eklenmesini desteklediğini ancak bu sistemlerin öğretmen değerlendirmesi kadar iyi olmadığını düşündüklerini göstermiştir. Öğrencilerin değerlendirme ve geri bildirimlere bakışı başarı seviyesine göre farklılaşmış; üst düzey başarı grubundaki öğrenciler YZ sistemlerinin "objektiflik" tarafını vurgularken, alt düzey başarı grubundaki öğrenciler "anlayışsızlık" eleştirisi yapmıştır. Araştırma sonuçları, YZ sistemlerinin tutarlılık, nesnellik ve hız gibi avantajlar sunduğunu ancak esneklik, bağlamsal anlama ve duygusal açıdan sınırlı kaldığını ortaya koymuştur. Bu bulgular doğrultusunda, YZ sistemleri ile öğretmenlerin birbirlerini tamamladıkları bütünleşik bir değerlendirme modelinin benimsenmesi önerilmektedir.

Anahtar Sözcükler: Yapay zeka sistemleri, matematik eğitimi, açık uçlu soruların değerlendirilmesi, öğretmen değerlendirmesi, psikometrik analiz, puanlayıcılar arası tutarlılık

ABSTRACT

A COMPARISON OF ARTIFICIAL INTELLIGENCE AND TEACHER SCORING IN THE ASSESSMENT OF OPEN-ENDED MATHEMATICS QUESTIONS

SENER, Serra Sezin

Master Degree, Department of Educational Sciences

Supervisor: Prof. Seher YALCIN

July 2025, xii + 101

The aim of this study is to analyze the place of systems supported by artificial intelligence (AI) in evaluation of mathematics questions through a comparative analysis with teacher assessments. A mixed method was adopted in the subject study; 217 7th grade students and their teachers from 5 different campuses of a private school (Ankara Oran, Ankara İncek, Antalya Muratpaşa, Antalya Manavgat, Sakarya) were studied. In the data collection process, the 10-item mathematics test of the 2nd semester's 1st common exam was used in 5 schools at the same time, and the exam papers of the students were evaluated by both the teachers and the AI system and feedback was given to the students.

Quantitative data were analyzed through reliability, item-total correlation, distinctiveness, and weighted Kappa agreement analyses, while qualitative data were collected through a questionnaire with 36 teachers, individual interviews with the mathematics teachers and assessment and evaluation experts of the school and focus group discussions with 30 students from different achievement levels determined by the exam results.

According to the outcomes of the study, the Cronbach's alpha reliability coefficient of the AI system was higher than the teacher assessment. The item-total correlations and distinctiveness indices also showed that the AI system was more consistent; furthermore, the reliability of both assessments was high. However, on the one hand evaluator correspondence analysis, high correspondence (.60 and above) was observed in algorithmic operations and formula applications, on the other hand low correspondence

(.40 and below) was found in questions requiring conceptual understanding and multi-stage problem solving.

The items with the largest score differences include mathematics topics such as multi-stage percentage calculations, real-life applications, and creating equations from verbal expressions.

The qualitative findings showed that the majority of teachers supported the integration of AI systems into education, but they did not think that these systems were as good as teacher assessment. The students' views on the subject were different based on their achievement levels. While the high-level achievement group considers the AI systems as “objective”, the low-level achievement group considers as “unsympathetic”.

The results revealed that the AI systems offer advantages such as consistency, objectivity and speed, but are limited in terms of flexibility, contextual understanding and emotionality. In line with these findings, it is recommended to adopt an integrated assessment model in which AI systems and teachers working in harmony.

Key Words: Artificial intelligence systems, mathematics education, assessment of open-ended questions, teacher evaluation, psychometric analysis, inter-rater reliability

ÖNSÖZ

Ölçme ve değerlendirme alanında, öğrencilerin bilgi ve beceri düzeylerini objektif bir şekilde ölçmek kritik öneme sahiptir; ancak geleneksel puanlama yöntemleri açık uçlu sorularda tutarlılık, nesnellik ve geri dönütlerin hızlı olması açısından çeşitli sınırlılıklar barındırmaktadır. Son yıllarda YZ teknolojilerindeki gelişmeler, otomatik puanlama sistemlerinin geliştirilmesini sağlamıştır. Bu sistemler sayesinde öğretmenlerin başarı sınavlarını değerlendirmeleri sürecindeki iş yükü azalırken; YZ sistemleri ile yapılan puanlamanın daha tutarlı ve objektif olması olasıdır.

Bu çalışmada, matematik dersi kapsamında hazırlanan açık uçlu soruların YZ destekli sistemlerin puanlamaları ile öğretmen puanlaması arasındaki farklılıklar incelenmiştir. Çalışmanın temel amacı, YZ sistemlerinin geleneksel öğretmen puanlamasına alternatif olup olamayacağını araştırmak; değerlendirme ve geri bildirim konusunda her iki puanlamanın güçlü ve zayıf yönlerini ortaya koymaktır. Elde edilen bulgular ile bu araştırmanın sonuçlarının eğitime dijital sistemlerin entegrasyonu sürecinin daha bilinçli bir şekilde yönetilmesine katkı sağlaması beklenmektedir.

TEŞEKKÜR

Bilgi birikimini benimle paylaşmaktan çekinmeyen, sabırla yol gösteren ve bana yalnızca öğrenci olarak değil bir birey olarak değer veren yaklaşımı için; akademik yolculuğumda önümü aydınlatan danışman hocam Prof. Dr. Seher YALÇIN'a sonsuz teşekkürlerimi sunuyorum.

Eleştiri ve önerileri ile çalışmamın daha da güçlenmesini sağlayan değerli jüri üyelerim Prof. Dr. Celal Deha DOĞAN ve Doç. Dr. Münevver İlgün DİBEK'e, gösterdikleri ilgi ve sundukları değerli katkılar için teşekkür ederim.

Bu süreçte en büyük destekçim olan, yüksek lisansı bitirmek dışında hiçbir sorumluluğum yokmuş gibi hissettiren canım eşim Kerim ŞENER'e ve yanında olmama ihtiyacı olduğu halde büyük bir anlayışla bekleyen biricik kızım Mavi'ye; anneme, babama ve kardeşime manevi desteklerinden ve hayallerimin peşinden koşmam için bana verdikleri güçten dolayı minnettarım.

Yüksek lisans yapmam için elimden tutan, desteğini her zaman hissettiğim değerli insan Dr. Levent OKUT'a; hem iş hayatımda hem eğitim hayatımda yanımda olduğu için teşekkür ederim.

Tez yazım sürecinde yanımda olan, stresli anlarımda beni motive eden ve bu zorlu süreci benimle paylaşan sevgili arkadaşlarım Melike BOZKURT, Sibel KAHRAMAN, Burcu BAYRAK, Pelin ÖÇAL, Dilek TOKHAY ve Sedef SERTTAŞ'a her kahve molası, her beyin fırtınası ve her kahkaha için sonsuz teşekkür ederim.

İÇİNDEKİLER

	Sayfa
ETİK İLKELERE UYGUNLUK BİLDİRİMİ.....	iii
ÖZET.....	iv
ABSTRACT.....	vi
ÖNSÖZ.....	viii
TEŞEKKÜR.....	ix
İÇİNDEKİLER.....	x
TABLolar VE GRAFİKLER DİZİNİ.....	xii
BÖLÜM 1.....	1
GİRİŞ.....	1
Problem.....	7
Amaç.....	8
Önem.....	8
Sınırlılıklar.....	9
BÖLÜM 2.....	10
KURAMSAL ÇERÇEVE.....	10
Eğitimde Değerlendirme.....	10
Geleneksel ve Tamamlayıcı Değerlendirme Yöntemleri.....	11
Ortaokul Matematik Müfredatında Değerlendirme.....	12
Açık Uçlu Sorular.....	13
Açık Uçlu Soruların Avantaj ve Dezavantajları.....	14
Puanlama Yöntemleri.....	15
YZ ile Otomatik Puanlama Yöntemleri.....	16
Rubrik Kullanımı, Önemi ve Türleri.....	17
Öğretmen Puanlamalarında Tutarlılık ve Güvenilirlik.....	19
Yapay Zekâ ve Eğitim.....	20
Yapay Zekâ ile Açık Uçlu Soruların Değerlendirilmesi ve Geri Bildirim.....	21
Öğretmen ve Yapay Zekâ Değerlendirmelerinin Karşılaştırılması.....	23
Güvenirlik ve Geçerlik Karşılaştırmaları.....	23
İlgili Araştırmalar.....	25
Ulusal Literatürdeki Çalışmalar.....	25
Uluslararası Literatürdeki Çalışmalar.....	32
BÖLÜM 3.....	37
YÖNTEM.....	37
Araştırmanın Modeli.....	37
Çalışma Grubu.....	37
Verilerin Toplanması.....	39
Veri Toplama Araçları.....	39
Veri Toplama Yöntemi.....	43
Verilerin Çözümlemesi.....	44
BÖLÜM 4.....	47
BULGULAR VE YORUMLAR.....	47

Matematik Yazılı Sınavının Güvenirlik ve Madde Özellikleri.....	47
Testin Güvenirliğine Dair Bulgular	47
Madde Güçlük İndeksleri	49
Madde Ayırt Edicilik İndeksleri	51
YZ ve Öğretmen Değerlendirmelerinin Karşılaştırmalı Analizi	52
Ağırlıklandırılmış Kappa Uyum Analizi Sonuçları	53
YZ ve Öğretmen Geri Bildirimlerinin Nitel Analizi	56
YZ Tabanlı Değerlendirme Sistemlerinin Eğitim Süreçlerine Entegrasyonu	60
BÖLÜM 5.....	71
SONUÇ VE ÖNERİLER	71
Sonuçlar	71
Öneriler	74
Uygulayıcılara Yönelik Öneriler.....	74
Politika Yapıcılara Yönelik Öneriler	75
Araştırmacılara Yönelik Öneriler.....	75
KAYNAKLAR.....	76
EKLER	81
EK 1. 2023-2024 Eğitim Öğretim Yılı 7. Sınıf 2. Dönem Matematik Yazılısı Soruları	82
EK 2. Dereceli Puanlama Anahtarı	84
EK 3. Başarı Seviyesine Göre Belirlenen Odak Grupların Öğretmen Puanları Tablosu	92
EK 4. Başarı Seviyesine Göre Belirlenen Odak Grupların YZ Puanları Tablosu	93
EK 5. Etik Kurul Raporu	94
EK 6. Okul Müdürleri Onam Formları	95
BENZERLİK BİLDİRİMİ	100
ÖZGEÇMİŞ.....	101

TABLolar VE GRAFİKLER DİZİNİ

Tablo 1. Katılımcı Bilgileri	39
Tablo 2. 7. Sınıf Matematik Dersi 2. Dönem 1. Yazılı Konu Soru Dağılım Tablosu	40
Tablo 3. Örnek Açık Uçlu Madde Dereceli Puanlama Anahtarı	41
Tablo 4. Kappa İstatistiğinin Yorumlanmasına İlişkin Değer Aralıkları	45
Tablo 5. Veri Çözümleme Yöntemleri	46
Tablo 6. Testin Güvenirlik Katsayıları	47
Tablo 7. Öğretmen ve YZ Puanlamasında Madde Toplam İstatistikleri	48
Tablo 8. Madde Güçlük İndeksleri (Kazanım Bazında)	50
Tablo 9. Madde Ayırt Edicilik İndeksleri ve Yorumları	51
Tablo 10. Değerlendiriciler Arası Madde Puanı Uyumu	53
Tablo 11. Değerlendiriciler Arası Tam Puan Uyumu	54
Tablo 12. Değerlendiriciler Arası Büyük Puan Farklılıkları	54
Grafik 1. Madde Bazında Puanlayıcı Uyumu Karşılaştırmaları	55
Tablo 13. Odak Grup Görüşmesi Geri Bildirim Analizi	56
Tablo 14. Öğretmen Anketinde Yer Alan Her Maddeye İlişkin Verilerin Yanıtların Dağılımına İlişkin Frekans ve Yüzdelere	61
Tablo 15. Öğrencilerle Yapılan Odak Grup Görüşmesi Entegrasyon Süreçleri Analizi	63
Tablo 16. Ölçme Uzmanları ve Matematik Öğretmenleri İle Yapılan Birebir Görüşmenin Analizi	68

BÖLÜM 1

GİRİŞ

İnsanları hayvanlardan ayıran en önemli özellik, bilgiyi diğerlerine aktarabilmesidir. Beyinlerinin büyüklüğü ve dil yetenekleri sayesinde medeniyet kurmuş ve sürekli gelişmişlerdir (Harari, 2015). İnsan beyni diğer hayvanlara göre daha büyük ve karmaşık olup, zeka, dil, kültür gibi üst düzey bilişsel yeteneklere sahip olmasını sağlamıştır (Bakırcı, 2011). Bu bilgi aktarımının sistemli hale gelmesi ile eğitim ortaya çıkmıştır. Eğitim, "bireyin davranışlarında kendi yaşantısı yoluyla istendik değişme meydana getirme süreci" olarak tanımlanmaktadır (Ertürk, 2013).

İnsanlar doğuştan çok az içgüdü ile geldiğinden, yaşam boyunca öğrenmek zorundadır (Selçuk, 2009). Öğrenme örtük bir psikolojik yapı olduğu için gözlemlenebilir ve ölçülebilir hale getirilmesi gerekmektedir. Bu durum eğitim sürecinin planlanması ve değerlendirilmesi için kritik önem taşımaktadır.

Öğrenme doğrudan gözlemlenemez; davranıştaki değişimler yoluyla ölçülür (Tekin, 1996). Bilgi, beceri ve davranışların edinilip edinilmediğinin ölçülmesi ve değerlendirilmesi gerekmektedir (Erkan vd., 2008). Tarihi süreçte eğitim 2005 yılına kadar öğretmen merkezli, bilgi aktarım odaklı ve sonuç odaklı değerlendirme anlayışı ile yürütülmüştür (Kutlu vd., 2017). Ancak ölçme ve değerlendirme süreçlerini eğitim alanında etkili biçimde kullanmak bir çok açıdan önem taşımaktadır. Özellikle öğretim hedeflerine hangi düzeyde erişildiğinin saptanması; öğretim yöntem ve materyallerinin amaca uygunluğun belirlenmesi yanında öğretim kararlarının alınabilmesi; öğretim sürecinin konu içeriği ve öğrenci nitelikleri ile uyumlu olmasının sağlanması gibi eğitim açısından önemli alanlarda etkili ölçme ve değerlendirme uygulamaları sayesinde ilerleme kaydedilmektedir. Öğretimsel kararların konuya ilişkin anlamlı ve güvenilir veriler üzerine kurulması, öğrencilerin öğretim süreci konusunda düzenli bilgilendirilmesi, başarısızlık yaşayan öğrenciler için çözüm yolları üretilmesi, başarılı öğrencilerin motivasyonunun artırılması ve öğrenme sürecindeki aksaklıkların giderilmesi de yine ölçme ve değerlendirmenin etkin kullanımı ile mümkün olmaktadır (Erkan vd., 2008).

2005 sonrasında öğretim programları değişmeye başlamış ve alt düzey düşünme gerektiren anlayıştan uzaklaşmıştır. Üst düzey düşünme becerilerini öne çıkartan bir anlayışa yönelmek, öğretim yöntem ve tekniklerinin de değişmesini; öğretmen merkezli olmaktan çıkıp öğrenci merkezli bir yapıya dönüşmesini gerektirmiştir. Eğitimin dönüşmesi ölçme ve değerlendirme yaklaşımlarının da dönüşmesini zorunlu kıldığı için sonuca odaklanan, sadece bilgilerin kazanılma durumunu ölçen anlayıştan uzaklaşmıştır. Öğrenilen bilgilerin olduğu gibi kullanılması yerine yeni durumlar veya gerçek yaşam örnekleri üzerinden, bilginin nasıl kullanılması gerektiğini ölçen bir yapıya dönüşmüştür (Kutlu vd., 2017).

Günümüzde eğitim sistemlerinin öncelikleri, kalite ve etkinliğin artırılması; bireyselleştirilmiş analizler ile öğrencilerin bireysel olarak ihtiyaçlarının karşılanması olarak karşımıza çıkmaktadır. Özellikle temel derslerde, üst düzey düşünme, problem çözme ve analitik becerilerin geliştirilmesi gibi süreçlere daha uygun olması nedeniyle açık uçlu sorulara yönelik değerlendirme yöntemleri önem kazanmaktadır. Ancak bu tip maddelerden oluşan testlerin; değerlendirmenin zaman alıcı olması, öznel yargılama ve puanlama olasılığının yüksek olması gibi olumsuz tarafları bulunmaktadır. Bu nedenle yapay zeka (YZ) teknolojilerinin bu süreçte bir yardımcı olarak kullanılması eğitimciler ve araştırmacılar için cazip hale gelmiştir. YZ tabanlı değerlendirme sistemlerinin, öğretmenlerle birlikte veya bağımsız olarak öğrenci çalışmalarını öznel ve etkin bir şekilde değerlendirme potansiyeline sahip olup olmadığı ortaya çıkartılıp eğitimde ölçme ve değerlendirmenin geleceğine katkıda bulunmak önem kazanmaktadır.

YZ destekli puanlama sistemleri ve bununla ilgili araştırmalar ülkemizde henüz yaygınlaşmış durumda değildir. Akdeniz ve Özdiç (2021) araştırmasında YZ'nin eğitimde kullanımını ve bu konuda Türkiye'de yapılan çalışmaları incelemişlerdir. Araştırmada, Ulusal Tez Merkezi ve Türkiye adresli dergiler taranarak elde edilen 37 çalışma, son yıllarda eğitimde YZ konusunda yapılan çalışmaların sayısında hızlı bir artış olduğunu göstermektedir. Çalışmanın bulgularına göre, 1999-2018 yılları arasında yapılan çalışmaların çoğunluğu tasarım ve geliştirme yöntemi ile gerçekleştirilmiştir, bu durum YZ alanında yeni teknolojilerin geliştirildiğini göstermektedir. YZ tekniklerinin en çok ele alındığı alan zeki öğretim sistemleri olarak belirtilmiştir. Bu sistemler, YZ tekniklerini kullanarak öğrencinin öğrenme sürecindeki performansını değerlendirir ve eğitim programını buna göre düzenleyebilir.

Çetin ve Aktaş'ın (2021) çalışmasında, YZ'nin eğitim alanına etkisi ve bu alanda nasıl kullanılabileceği üzerine inceleme yapılmıştır. YZ'nin eğitim alanındaki

potansiyeli, faydaları ve olası olumsuz etkileri belirtilmiş; ayrıca bu teknolojinin mevcut eğitim sistemine eklemlenmesi ve geleceği araştırılmıştır. Eğitime henüz tam anlamıyla eklemlenmemiş olsa da, yakın gelecekte eğitimin temel paydaşlarını etkileyeceği öngörülmüştür. Makalede YZ'nin eğitim alanındaki geleceği “Sınıf Yönetiminde Yapay Zeka” ve “Okul Yönetiminde Yapay Zeka” şeklinde iki ana başlıkta değerlendirilmiştir. YZ'nin öğretmenin ve okul müdürlerinin yerini alabileceği; öğrencilerle ilişkiler, ölçme-değerlendirme, müfredat ve ders planlaması gibi konularda önemli faydalar sağlayabileceği ancak öğretmenlerin pedagojik ve duygusal yaklaşımlarını tam olarak taklit edemeyeceği için öğretmenin yerini tam anlamıyla alamayacağı, yardımcı bir rolde faydalı olabileceği vurgulanmıştır. YZ'nin insan duyguları konusunda bazı eksiklikler yaşayacağı düşünülmektedir. İdari koordinasyon ve kontrol gibi işleri otomatikleştirebilirken, insani ilişkiler ve karar verme gibi konularda insan liderlerin yerini tutamayacak denilmiştir.

Ülkemizdeki çalışmalar çoğunlukla YZ'nin eğitimde öğrenmeye katkısı, eğitsel araç olarak ya da idari olarak kullanımı üzerine yoğunlaşmış ya da Çelik ve Çoban'ın (2023) çalışmasında olduğu gibi açık uçlu maddelere YZ'nin verdiği yanıtlar değerlendirilmiştir. Araştırmada, iki farklı sohbet robotunun üniversite düzeyinde, Bloom'un bilişsel alan taksonomisine uygun hazırlanmış yüzey gerilimi sorularına verdiği yanıtların değerlendirilmesi üzerine çalışılmıştır. Altı açık uçlu soruya verilen YZ yanıtları uzmanlar tarafından değerlendirilmiş ve yanıtların çoğunlukla (%66,7) net olduğu saptanmıştır. Satranç oyununda hatta daha sonrasında uzun yıllar boyunca bilgisayarlaştırılmaya direnen Go oyununda insanlara karşı mutlak başarılar alan bilgisayar programlarına hayret edilen 2016 yılından günümüze kadar en çok merak edilen konu hep “YZ da insan gibi düşünebilir mi?” olmuştur (Say, 2018). Bu amaçla genel olarak araştırmalar sorulan sorulara YZ'nin verdiği yanıtları değerlendirmek üzerine yoğunlaşmıştır. Ülkemizdeki araştırmalarda insanlar tarafından yanıtlanmış açık uçlu maddelerden oluşan testlerin YZ'ye puanlatılması çalışmasına 2025 yılına kadar rastlanmamaktadır.

Yılmaz (2025), doktora tez çalışmasında Türkçe dersine ait açık uçlu maddelerin otomatik puanlanmasında çeşitli makine öğrenmesi ve doğal dil işleme yöntemlerini karşılaştırmıştır. Araştırmada denetimli makine öğrenmesi modellerinden Naive Bayes, Lojistik Regresyon, Karar Ağacı, Rastgele Orman ve Gradyan Artırma algoritmaları ile dönüştürücü tabanlı BERT modeli kullanılmıştır. Python programlama dili ve çeşitli kütüphanelerin kullanıldığı çalışmada, modellerin performansı Ağırlıklandırılmış Kappa

(Quadratic Weighted Kappa - QWK) ile değerlendirilmiştir. Sonuçlar, BERT modelinin en yüksek makine-insan uyumunu (0,72-0,93) sağladığını, bunu Gradyan Artırma algoritmasının (0,67-0,90) takip ettiğini göstermiştir. Naive Bayes ise tüm maddeler için en düşük performans gösteren algoritma olmuştur (0,44-0,78). Bu bulgular, özellikle Türkçe dilindeki açık uçlu maddelerin otomatik değerlendirilmesinde BERT gibi gelişmiş dil modellerinin daha etkili olduğunu ortaya koymaktadır.

Yurt dışı alan yazına bakıldığında zaman 2020 ve sonrasında metin madenciliği ve büyük dil modelleri (LLM) ile matematik alanındaki açık uçlu soruların puanlanarak insan puanlayıcıların sonuçları ile karşılaştırıldığı beş çalışma olduğu görülmektedir. Bu çalışmalar kronolojik olarak aşağıda sunulmuştur.

McNichols ve diğerleri (2023), henüz hakemli bir dergide yayınlanmamış, Cornell Üniversitesi makale arşivinde (arXiv) yer alan çalışmada öğrencilerin cebirsel problemleri çözerken yaptıkları hataların sınıflandırılmasında geleneksel hata sınıflandırma sistemleri yerine, daha esnek ve genelleştirilebilir bir yaklaşım sunabilmek için LLM'lerin kullanılabilirliği incelemiştir. Geleneksel hata sınıflandırma sistemleri, alan uzmanları tarafından geliştirilmiş el yapımı kurallara dayandığından sistemlerin genelleştirme kapasitesi sınırlı olmaktadır. Mevcut YZ sistemlerinin yöntemleri ise matematik ifadelerinin hiyerarşik yapılarını görselleştiren sözdizimi ağaçlarına (Syntax trees) dönüştürülmesini gerektirmekte, ancak öğrenci yanıtları her zaman bu şekilde ifade edilememektedir. Bu nedenle çalışmada önceden eğitilmiş LLM'lerin (BERT, GPT-2, MathBERT, XLM-RoBERTa) cebir hatası sınıflandırması için ince ayarlama (Fine-tuning) yaklaşımı ile yeni göreve uyarlanması sağlanmıştır. Toplamda 130.823 etkileşim, 9.590 yanıt ve 95 öğrenci verisi kullanılarak, hata tespit edilen 5.744 adım üzerinde çalışılmıştır.

BERT modeli %80.68 doğruluk oranıyla en iyi performansı göstermiş; öğrenci eylem bilgilerinin modele dahil edilmesi ile önemli performans artışı sağlanmıştır (%82.02'den %85.90'a). Bunun yanında matematik alanına özgü ön eğitim, ilk eğitim dönemlerinde küçük iyileştirmeler sağlamış ancak genel performansta anlamlı bir değişiklik yaratmamıştır. Çalışma, LLM tabanlı yöntemlerin matematik eğitiminde hata sınıflandırması için etkili bir çözüm sunduğunu göstererek eğitim alanında öğrenci hatalarının otomatik tespiti ve sınıflandırılması konusunda önemli bir katkı sağlamakta ve eğitim teknolojilerinin geliştirilmesi açısından değerli bulgular sunmaktadır (McNichols vd., 2023).

McNichols, Lee ve diğerlerinin (2024) Eğitimde Veri Madenciliği (Educational Data Mining-EDM Conference) bildirisi olarak yayımlanmış olan çalışmada, LLM'lerin akıllı öğretim sistemlerinde (ITS) kullanılan önceden hazırlanmış mesaj metinlerini kullanarak öğrencilere otomatik geri bildirim sağlama sürecini taklit edebilme kapasitesi üzerine araştırma yapılmıştır. ITS'ler, tanımlanmış hata türleri konusunda, insan uzmanlar tarafından hazırlanan şablonlar ile geri bildirim verdiği için yalnızca önceden öngörülmüş hatalar için geri bildirim verebilmektedir. Bu nedenle her yeni soru türü için yeni şablonlar geliştirmek gerekmektedir. LLM'lerin, var olan ITS sistemlerinin bu mekanizmalarını otomatik olarak taklit edebilmeleri için ince ayarlama (Fine-tuning), bağlam içi öğrenme (In-Context Learning) ve her iki yöntemin birleşimini içeren hibrit yaklaşım ile Mistral-7B ve GPT-3.5-turbo modelleri test edilmiştir. Aynı hata türündeki farklı yanıtlar, görülmemiş sorular için geri bildirimler ve tamamen yeni hata türleri şeklinde üç farklı veri bölünmesi stratejisi kullanılan çalışmada LLM'lerin şablon yapısını başarıyla öğrenebildiği, yeni sorular için kabul edilebilir geri bildirim üretebildiği ancak yeni hata türlerini anlayamadığı sonuçlarına ulaşılmıştır. Modellerin karşılaştırılmasında ise GPT-3.5 modelinin tüm durumlarda Mistral-7B'den üstün performans sergilediği görülmüştür.

Sonuç olarak LLM'lerin ITS sistemlerinin geri bildirim verme şeklini öğrenebildiği ama matematik hatalarını gerçekten anlamadığı ve hataları genellemediği için geri bildirimlerin eğitim sırasında görülen hata türlerine benzeyen hata durumları ile sınırlı olduğu görülmüştür. Bu çalışma, LLM'lerin ITS geri bildirim mekanizmalarını taklit etmek konusunda umut vaat ettiğini ancak gerçek matematik bilimi ve teorisinde sınırlı kalması nedeniyle insan eğitimcilerin yerini almak değil, onları destekleyecek araçlar olarak kullanılması gerekliliğini göstermiştir (McNichols vd., 2024).

Baral ve diğerleri (2025), çalışmalarında, matematik eğitiminde açık uçlu sorulara verilen öğrenci yanıtları için otomatik geri bildirim üretmede LLM'lerin etkinliğini araştırmak amacıyla üç farklı model karşılaştırılmıştır (SBERT-Canberra, GOAT ve GPT-4). Geleneksel geri bildirim sistemlerinde yaşanan; sınırlı hata türlerini kapsamaması, insan uzmanların çalışması gerekliliği ve ölçeklenebilirlik problemi yaşanması sorunlarının aşılması amacıyla yapılan çalışmada açık uçlu soruların önemi üzerinde durularak öğretmen puanlamalarını en iyi tahmin eden yüksek performanslı modelin hangisi olduğu sorusuna yanıt aranmıştır. “Illustrative Mathematics” müfredatından 50 açık uçlu matematik sorusuna 100 öğrencinin verdiği yanıtlar kullanılmıştır.

SBERT-Canberra modeli ince ayar yapılmadan, geçmiş benzer yanıtları bularak puanlama yaparken GOAT, Mistral 7B'den geliştirilmiş ve 4.000 örnekle eğitilmiştir. GPT-4 ise özel eğitim almadan sadece rubrik bilgisiyle test edilmiştir. Temel olarak GOAT modelinin tüm puanlama metriklerinde en iyi performansı gösterdiği ancak geri bildirim kalitesinde GPT-4'ün açık ara önde olduğu; ince ayarlı modellerin (GOAT, SBERT) puanlamayı başarılı yaparken gelişmiş dil üretme becerisi nedeniyle GPT-4'ün geri bildirim kalitesinin yüksek olduğu ortaya konmuştur. İnce ayarlı modellerin başarısı kullanılan eğitim verisi kalitesine bağlı olduğu için puanlamanın bu LLM'ler ile yapılması; ancak GPT-4 gibi güçlü modellerin geri bildirim üretiminde üstün olmaları nedeniyle her modelin güçlü yönlerini birleştiren hibrit sistemlerin olması gerektiği sonucuna varılmıştır (Baral vd., 2025).

Wang ve diğerleri (2025), çalışmalarında üç LLM'nin (GPT-4o, DeepSeek-V3, Gemini-2.0) farklı karmaşıklık seviyelerindeki matematik problemlerini çözme yetenekleri değerlendirilmiştir. Değerlendirme, Yapılandırılmış Düşünce Zinciri (SCoT) çerçevesi kullanılarak 5.000 temel seviye aritmetik problemi, 500 orta-ileri seviye matematiksel muhakeme problemi ve 130 lisansüstü seviye gerçek sınav soruları kullanılarak beş temel boyutta yapılmıştır (son yanıtın doğruluğu, çözüm adımlarının bütünlüğü, her adımın geçerliliği, ara hesaplamaların kesinliği, problem anlayışının derinliği).

Araştırma sonucunda; genel matematik problemlerinde en başarılı modelin Gemini-2.0; karmaşık ve istatistiksel problemlerde en güçlü modelin GPT-4o ve hesaplamalı görevlerde DeepSeek-V3'ün diğer modellerle yarışabilir seviyede olduğu görülmüştür. Tüm modellerin soru zorlukları arttıkça performans kaybı yaşadığı ve soyut matematik sorularında zorlandıkları görülmüş; çok adımlı algortimalarda yetersizlik, mantıksal tutarlılık sorunları, formal teoremleri ispatlamada eksiklik gibi sınırlılıklar olduğu belirtilmiştir. Bu çalışmada da diğer araştırmalarda olduğu gibi farklı modellerin güçlü yönlerinin birleştirildiği hibrit sistemlerin kullanılması önerilmektedir (Wang vd., 2025).

Bu bağlamda, açık uçlu maddelerin değerlendirilmesinde YZ tabanlı sistemlerin potansiyeli dikkat çekmektedir ancak ulusal ve uluslararası literatür incelendiğinde, yapılan çalışmaların çoğunlukla öğrenci yanıtlarının analizine odaklandığı, öğretmen ve YZ puanlamalarının tutarlılığını inceleyen çalışmaların ise sınırlı olduğu görülmektedir. Bu durum, ölçme-değerlendirme süreçlerinin güvenilirliği açısından önemli bir araştırma ihtiyacı doğurmaktadır.

Problem

Eğitimde üst düzey düşünme becerileri ve problem çözme becerilerinin ölçülmesi sürecinde performans değerlendirme ile birlikte sınıf içi ölçmelerde özellikle açık uçlu soru kullanımı zorunlu hale gelmiştir. Bununla birlikte 2023 - 2024 eğitim öğretim yılı başında Milli Eğitim Bakanlığı (MEB) tarafından alınan bir karar ile sınıf içi ölçmelerde uygulanacak yazılı yoklamaların yalnızca yanıtları öğrenciler tarafından oluşturulacak açık uçlu ya da kısa yanıtlı sorulardan oluşabileceği; sene başında bakanlık tarafından ilan edilerek gerekli görülen derslerde tüm ülke genelinde ortak yazılılar yapılabileceği; bunun yanında yine sene başında ilan edilmek şartıyla il müdürlükleri ya da ilçe müdürlükleri tarafından da il ya da ilçe genelinde ortak yazılılar yapılabileceği duyurulmuş ve Resmi Gazete’de yayımlanmıştır. Okul içinde yapılacak ortak yazılılarda kullanılacak soruların ders öğretmenlerince; ülke, il ya da ilçe genelinde yapılacak ortak yazılıların ise bakanlık ya da müdürlüklerce hazırlanacağı bildirilmiştir. Bu kapsamda ilk örnek olarak 2023 - 2024 eğitim öğretim yılında 6. sınıf seviyesinde Türkçe ve matematik ile 9. sınıf seviyesinde Türk dili ve edebiyatı ve matematik derslerinin her iki dönemdeki ilk yazılıları ülke genelinde ortak yazılı olarak uygulanmıştır. Ülke geneli ortak yazılılarda; ilk dönemin ilk yazılıları çoktan seçmeli maddelerden oluşurken ikinci dönemin ilk yazılılarında Türkçe dersinin on altı maddesinden iki tanesi, matematik yazılısının ise tamamı açık uçlu maddelerden oluşturulmuştur. Maddelerin puanlanması için ülke genelinde gönüllü puanlayıcılara ulaşılmış ve öğrencilerin yazılı kağıtları dijital ortama aktarılarak gönüllüler tarafından puanlanması süreci başlatılmıştır. Ancak bu işlem hem zaman almakta hem de puanlamanın tarafsızlığının sağlanamaması gibi sorunlar ortaya çıkabilmektedir. Bu bağlamda, MEB tarafından sadece matematik alanındaki soruların tamamında açık uçlu soruların kullanılmış olması nedeniyle bu çalışma da matematik alanıyla sınırlandırılmıştır. Elde edilecek bulguların MEB uygulamalarına katkı sağlaması umulmaktadır.

Açık uçlu sorulardan oluşan bir sınavda, puanlamaların, test istatistiklerinin ve sonuç raporlarının oluşturulması ve her öğrenciye özelleştirilmiş geri bildirim sağlayan karnelerin hazırlanması sürecinde, insan puanlayıcılarına düşen yükün fazlalığı, bu konularda YZ kullanımına yönelik yapılan çalışmaların azlığı ve MEB tarafından alınan karar doğrultusunda tüm ülke genelinde puanlanması gereken açık uçlu testler karşısında insan puanlayıcıların yetersiz kalabileceği ve bu duruma alternatif yolların ortaya konulmasına yönelik ihtiyaç nedenleriyle bu çalışmanın yapılmasına gerek görülmüştür.

Amaç

Bu çalışmanın amacı, sınıf içi ölçmelerde kullanılan açık uçlu maddelerden oluşan bir matematik testinde, YZ ve öğretmen değerlendirmelerinin etkinliğini ve güvenilirliğini karşılaştırmak; yöntemlerin test istatistikleri ve sonuç raporları arasındaki tutarlılıkları ve farklılıklarını incelemek ve her öğrenciye özelleştirilmiş geri bildirim sağlayan karnelerin hazırlanmasında YZ ve öğretmen değerlendirmesi yöntemlerinin öğrenci başarısı ve motivasyonu üzerindeki etkilerini analiz etmektir.

Bu çalışma, YZ tabanlı değerlendirme sistemlerinin eğitim süreçlerine adapte edilmesini ve bu süreçte karşılaşılan zorlukları inceleyerek, ölçme ve değerlendirmede YZ kullanımının önemini ortaya koymayı amaçlamaktadır. Eğitimcilerin yanında politika yapıcıların ve teknoloji geliştiricilerin de öğrenme süreçlerini desteklemek için YZ tabanlı araçları daha etkili bir şekilde nasıl kullanabileceklerine dair stratejiler geliştirmelerine yardımcı olmayı hedeflemektedir. Bu kapsamda yanıt aranan sorular aşağıda sunulmuştur:

1. YZ ve öğretmen puanlamalarının arasında tutarlılık ve objektiflik düzeyleri nasıldır?
2. Öğrencilere verilen bireyselleştirilmiş geri bildirimlerde, YZ'nin ve öğretmenin sunduğu geri bildirimler nasıl farklılık ve benzerlik göstermektedir?
3. YZ tabanlı değerlendirme sistemleri eğitim ortamlarında hangi yollarla uygulanmakta ve hangi amaçlarla kullanılmaktadır?
4. Öğretmenler ve öğrenciler, YZ tabanlı değerlendirme sistemlerinin eğitim süreçlerine entegrasyonunu nasıl değerlendirmektedir?
5. YZ tabanlı değerlendirme sistemlerinin entegrasyonu sürecinde hangi teknik ve pedagojik zorluklarla karşılaşmaktadır?

Önem

Bu araştırma, eğitimde YZ kullanımının etkinliğini ve güvenilirliğini ortaya koyarak, özellikle açık uçlu soruların değerlendirilmesi sürecinde öğretmenlerin karşılaştığı zorluklara yenilikçi bir çözüm sunmayı amaçlamaktadır. Açık uçlu sorular, öğrencilerin eleştirel düşünme, problem çözme ve yaratıcı düşünme becerilerini ölçmede önemli bir yere sahiptir; ancak, bu tür soruların değerlendirilmesi taraflı olabilir ve fazla

zaman harcanmasını gerektirebilir. YZ tabanlı sistemlerin kullanımı, değerlendirme sürecinde objektiflik ve verimliliği artırabilir, öğretmenlerin iş yükünü azaltabilir ve öğrencilere zamanında ve kişiselleştirilmiş geri bildirim sağlayabilir. Bu çalışma, YZ ve insan değerlendirmelerinin karşılaştırmalı analizini yaparak, eğitim teknolojilerinin gelişimine katkıda bulunmayı ve eğitimciler için yol gösterici olmayı hedeflemektedir. YZ destekli değerlendirmenin etkin kullanımının, açık uçlu testlerin hızlı puanlanması ve anında geri bildirim sağlaması sayesinde yapılan ölçmenin etkinliğini artıracak ve öğrencilerin hızlı dönütler alması ile motivasyonlarının artmasına yardımcı olacağı düşünülmektedir.

Bu çalışmanın, MEB tarafından 2023 - 2024 yılı başında alınan kararla tüm sınıf seviyelerinde sınıf içi ölçmelerde kullanılacak açık uçlu ya da kısa yanıtlı soruların değerlendirilmesinde öğretmenlere; Türkiye geneli ortak yazılılarda açık uçlu soruların değerlendirilmesinde MEB ölçme ve değerlendirme birimi ve ilgili tüm birimlerine ve ulusal sınavlarda açık uçlu soru yazımında, puanlamasında ve değerlendirilmesinde ÖSYM'ye katkı sağlayacağı düşünülmektedir. Ayrıca puanlayıcı güvenilirliğine yönelik yapılacak bu çalışmanın ölçme ve değerlendirme alanına da katkı sağlayacağı umulmaktadır.

Sınırlılıklar

Çalışmanın nicel kısmında toplam 217 öğrenciden, nitel kısmında ise 36 öğretmen ve üç farklı başarı grubundan 30 öğrenciden oluşan örneklem, özel okul öğrencilerini ve öğretmenlerini kapsamaması nedeniyle Türkiye genelindeki sosyoekonomik ve kültürel çeşitliliği tam olarak yansıtmayabilir.

Her okuldan yazılı sınava katılmış olan öğrencilerin puanlamasını o okulda görev yapan öğretmen puanlamıştır. Her öğretmenin farklı öğrencileri puanlamış olması YZ ile uyum konusundaki sınırlılığı oluşturmaktadır.

Araştırma, matematik dersinin yalnızca 7. sınıf 2. dönem MEB yazılı senaryosundaki kazanımları ile sınırlıdır. Bu senaryo çoğunlukla algoritmik işlemler ve formül uygulamalarını içermekte olup, matematik eğitiminin ispat yapma, matematiksel modelleme gibi diğer önemli boyutlarını kapsamamaktadır.

Araştırma, YZ algoritmalarından *Claude (Sonnet)* ile sınırlıdır.

BÖLÜM 2

KURAMSAL ÇERÇEVE

Bu bölümde, eğitimde değerlendirme yaklaşımları ve yöntemleri, ortaokul 7. sınıf matematik dersinde sınıf içi ölçmelerde kullanılan açık uçlu soruların değerlendirilmesi, açık uçlu soruların özellikleri ve puanlama yöntemleri incelenerek eğitimde değerlendirme ile ilgili temel kavramlara yer verilmiştir. Bununla birlikte YZ tabanlı sistemlerin eğitimde kullanımı ile açık uçlu soruların değerlendirilmesindeki rolü ve öğretmen puanlamalarıyla YZ puanlamalarının karşılaştırılmasına yönelik literatür taraması sunulmuştur. Ulusal ve uluslararası literatürde yer alan çalışmalar incelenmiş; bu sayede araştırmanın kuramsal temelleri ortaya konulmuştur.

Eğitimde Değerlendirme

Değerlendirme “ölçme sonuçlarının bir ölçütle karşılaştırılması sonucunda bir değer yargısına (karara) ulaşma işi”dir (Turgut, 1990). Bu tanımda üç temel öge bulunmaktadır; “ölçme sonucu, ölçüt ve karar”. Bu tanımdan yola çıkarak gerçek bir değerlendirme için yapılması gerekenler belirlenebilir. Öncelikle değerlendirilecek özellik ile ilgili ölçüt / ölçütler saptanır, gözlemler yapılır ve gözlem sonuçları ölçütlerle karşılaştırılarak karara varılır. Okullarda sınıf içi ölçme ve değerlendirme uygulamalarının temel amacı öğrencilere öğrenme durumları ile ilgili geri bildirim vermektedir. Eğitim sisteminin tüm paydaşları ile ilgili verilecek kararların doğru ve yerinde olması açısından değerlendirmenin önemi daha da çok anlaşılmaktadır. Bu bağlamda değerlendirme, ölçmeyi de içine alan; öğrenci, öğretmen, eğitim programı ve okullar ile ilgili verilecek kararları önemli ölçüde etkileyen geniş kapsamlı bir kavramdır (Bahar vd., 2008; Baykul ve Turgut, 2019).

Değerlendirmenin, eğitim sistemindeki kullanılış amaçlarına bakıldığında, üç farklı türde karşımıza çıkmaktadır. Tanılayıcı (diagnostic assessment), süreç (formative assessment) ve düzey belirleyici değerlendirme (Summative assessment) (Bahar vd., 2008). Tanılayıcı değerlendirme, eğitim sürecinin başında, eğitimi alacak kişilerin var olan durumunu öğrenmek için kullanılan araçlardır. Bu araçlar, tıpkı bir doktorun kan

testi gibi teşhis koymak için gerekli verileri elde etmeye yarar. Testler doğası gereği tanısaldır ve bir öğrencinin belirli bir kriter veya standartla ilgili performansını gösterir. (Popham, 2000). Süreç değerlendirme, öğrenme sürecini ve öğrenmeleri izlemeye yönelik değerlendirmedir. Öğrencilerin, süreçteki her üniteye ayrı ayrı öğrenme eksikliklerini belirlemek ve onların bu eksikliklerine yol açan güçlüklerin ne olduğunu anlamak için yapılır (Özçelik, 2016).

Düzey belirleyici değerlendirme, bir sınav katılımcısının bilgi ve becerilerinin genellikle bir öğrenme programının tamamlanmasından sonra gerçekleştirilen değerlendirmesidir (American Educational Research Association [AERA] vd., 2014). Öğrenme düzeyini belirleme amacı güder. Sonuçlarından elde edilen veriler ile eğitim sürecinde planlanmış olan hedefler doğrultusunda ne kadar ilerleme sağlandığı ortaya konur (Özçelik, 2016). Sınıf içi ölçmelerde öğretmen tarafından hazırlanarak, öğrencilere her dönem için not verilmesi amacı için yapılan yazılı yoklamalar da düzey belirleyici değerlendirme türündedir.

Geleneksel ve Tamamlayıcı Değerlendirme Yöntemleri

Daha önce de bahsedildiği gibi ölçme için gözlem yapılması gereklidir, gözlemler duyu organları aracılığı ile yapılmaktadır. Ancak eğitim ve psikoloji gibi davranış bilimlerinde ölçme konusu olan nitelikler doğrudan gözlenemediğinden, duyu organlarımızla gözlemleyemeyiz ve ölçmek için bir araç kullanmamız gerekir (Baykul ve Turgut, 2019). Eğitimle kazandırılmak istenen davranışın kazanılıp kazanılmadığını anlayabilmek amacı ile geliştirilmiş pek çok araç ve yöntem vardır. Ancak araç ve yöntemlerin hiçbiri mükemmel değildir, eksikleri ve geliştirilmesi gerekli noktaları vardır (Özçelik, 2013).

Eğitimde kullanılan ölçme araçları geleneksel ve tamamlayıcı olarak ayrılır. Performansa dayalı ölçme araçları (proje, sunum, drama, öz değerlendirme, akran değerlendirmesi, kontrol listesi vb) geleneksel yöntemlere ek olarak sunulduğu için tamamlayıcı araçlar olarak adlandırılmıştır. Geleneksel yöntemler ise yazılı yoklama, sözlü yoklama, kısa cevaplı sorular, kompozisyon, doğru-yanlış soruları vb. gibi öğretmen tarafından belli bir amaç için oluşturulmuş sorunun yanıtının öğrenci tarafından hatırlanarak tamamen baştan oluşturulması (açık uçlu yoklamalar) ya da var olan seçenekler içinden doğru olanın seçilmesi (çoktan seçmeli) şeklinde karşımıza çıkmaktadır.

Tamamlayıcı değerlendirme için kullanılan ölçme araçları öğrenmenin ölçülmesi ve değerlendirilme sürecinde öğrencinin etkin bir şekilde yer almasını hedefleyen uygulamalardır. Yapılandırmacı yaklaşımın olduğu eğitim sistemlerinde uygulanabilir. Bilginin önceki bilgilerle birleştirilip kullanılmasının ve arada bağlantı kurulabilmesinin önemli olduğu bir yaklaşımdır. Bu tür yöntemlerde amaç bilginin sadece hatırlanması değil yeni durumlar karşısında uygulanabilmesi ve açıklanabilmesidir; yani ürüne ve sürece dayalıdır (Özdemir, 2010).

Okullarda başarıyı değerlendirmek için kullanılan değerlendirme yöntemleri ise çoğunlukla geleneksel yöntemlerdir. Geleneksel değerlendirme, öğrencinin içinde bulunduğu örneklemden farkını ya da bu örneklemlerle olan benzerliğini belirlemek için gruptan elde edilen puanlar ile kişilerin puanlarının ayrı ayrı karşılaştırılarak incelenmesidir. Bilişsel yetenekler ve akademik başarının standart testleri ve birçok derecelendirme ölçeği bu şekilde yapılandırılmıştır (Evans vd., 2004).

Ortaokul Matematik Müfredatında Değerlendirme

Milli Eğitim Bakanlığı (2024), Ortaokul Matematik Dersi Öğretim Programı'nda, öğrenmeyi destekleyerek öğrencilere düzenli geri bildirim verilmesini sağlayacak bir yaklaşım benimsediğini belirtmiştir. Bu yaklaşımda *“öğrencilerin bilgi ve becerilerinin gelişimini izleme ve değerlendirmenin yanı sıra matematiğe yönelik eğilimlerinin, sosyal-duygusal öğrenme becerilerinin, okuryazarlık becerilerinin ve değerlerinin gelişiminin gözlenmesi”*nin de önemini vurgulamıştır. Tamamlayıcı (formatif) ölçme araçları kullanılarak öğrencilere bilgi düzeyleri, eksiklikleri veya kavram yanlışları hakkında geri bildirim verilmesi amaçlanmıştır.

Bunun yanında bakanlığın yayınladığı 9 Eylül tarihli yönetmelik ile ölçme ve değerlendirme yaklaşımı net bir şekilde açıklanmıştır. Yönetmeliğin üçüncü bölüm f bendinde *“Okullarda yapılacak ortak yazılı sınavların soruları konu soru dağılım tablosuna göre hazırlanır. Konu soru dağılım tablosu il sınıf/alan zümreleri ve Ölçme Değerlendirme Merkezi Müdürlüğü ile birlikte oluşturulur.”* diyerek yazılı sorularının oluşturulurken kapsam geçerliği açısından tabloların kullanılması gerektiği ve g bendinde *“Ülke ya da il/ilçe genelinde yapılacak ortak yazılı sınavlar hariç, okullarda yapılan tüm sınavlar cevaplarını öğrencilerin oluşturduğu ve farklı bilişsel düzeyde kazanımları ölçen maddelerden oluşan yazılı yoklama şeklinde yapılır.”* diyerek yazılı sorularının

tamamının geleneksel yöntemler kullanılarak açık uçlu sorular ya da kısa yanıtlı sorulardan oluşabileceği belirtilmiştir (Resmî Gazete, 2023).

Açık Uçlu Sorular

Başarı testleri belirli bir ders veya programdan bireylerin ne kadar yararlandıklarını göstermek için geliştirilen testlerdir; aynı zamanda bireylerin gelecekteki başarılarını yordamak amacıyla da kullanılır. Yetenek, kişilik ilgi ve tutum gibi diğer testler arasında en çok kullanılan test çeşidi başarı testleridir. Öğretmen yapımı testler ve standart testler olarak ikiye ayrılırlar. 19. yüzyıl ortalarına kadar okullarda başarının ölçülmesinde sözlü sınavlar kullanılmıştır. Sözlü sınavların subjektifliğinin eleştirilmesi ve herkese eşit zamanda, daha çok sayıda ve eşit sorular sorulmasının sağlanması düşüncesi ile Horace Mann tarafından yazılı sınavlar ortaya atılmıştır. Daha sonra objektif testlerin oluşturulması çalışmaları devam etmiş ve bu alana en büyük katkı; başarı testlerinin geliştirilmesi, hazırlanması ve sonuçlarının değerlendirilmesinde kullanılacak istatistiksel yöntemlerin belirlenmesi ile Thorndike tarafından yapılmıştır (Özgüven, 2007).

Başarı testinden elde edeceğimiz puanların neyi değerlendirmek için kullanılacağı belirlenmesi gerekir. Puanlarla süreç izlemeye yönelik bir sonuç elde edip öğrenciye geri bildirim mi verilecek, başarı sırası oluşturulup bir seçme - yerleştirme mi yapılacak, var olan bilgiler sınanıp ders programı mı oluşturulacak gibi kararlar testte yer alacak maddelerden elde edilecek puanlar için kullanılacak istatistik yöntemlerin seçilmesine kadar tüm alanlar için önemlidir (Kutlu vd., 2017). Okullarda puan verilmek amacı ile uygulanan yazılı yoklamalarda, Milli Eğitim Bakanlığının 2023 yılında aldığı kararla, sadece açık uçlu ya da kısa yanıtlı maddeler kullanılabileceği duyurulmuştur.

Açık uçlu maddeler, öğrencinin düşünmesi, yanıtı organize etmesi ve yapılandırarak yanıtlaması gereken madde türüdür. Analiz, sentez ve değerlendirme basamaklarında ölçme yapılabilir ve böylece ileri düzey davranışlar da ölçülebilir. Öğrencilerin yazıları ve imla kurallarına ne kadar uydukları gözlemlenebilir. Açık uçlu maddelerin puanlanmasında öğretmen tarafından yanlılık yapılabilir. Bu nedenle ayrıntılı / dereceli puanlama anahtarı oluşturulması ve birden çok puanlayıcı kullanılması gerekir. Hazırlaması ve uygulaması kolaydır ancak öğrenci sayısı az olduğunda kullanıma uygundur çünkü hazırlaması kolay olsa da puanlanma süreci uzundur (Kutlu vd., 2017).

Kısa yanıtlı maddeler; tek sözcük, rakam, tarih ... ya da bir iki cümle ile yanıtlanabilen maddelerdir. Yanıtı öğrenci tarafından hatırlanıp yazılan maddelerden

oluşur. Açık uçlu maddelere yakındır ancak istenen yanıtların kısa ve net olması nedeniyle objektiftir, bu nedenle puanlama hatası az olur. Çoktan seçmeli testlere göre daha üst düzey bir zihinsel etkinlik ölçer çünkü sorunun yanıtını öğrencinin hatırlaması gerekir. Maddelerin hazırlanması kolaydır. Yanıtlamak kısa sürdüğü için kısa zamanda daha çok soru sorulur (Tekin, 2010).

Değerlendirmenin amacı sadece not vermek değildir; öğrencileri bütün yönleriyle tanımak ve onların tüm gelişimi hakkında bir karara varmak asıl amaçtır. Zekâ, yetenek, ilgi, tavır ve bir çok kişilik özelliği her bireyde farklıdır ve bu nitelikler birbirlerini etkiler. Tüm bu özelliklerin gözlemlenmesi, örtük özellikler olmaları nedeniyle dolaylı olmaktadır; bu da ölçme işinin karmaşık ve zor olmasına yol açar. Bu nedenle eğitimde bir çok ölçme araç ve teknikleri geliştirilmiştir (Gümüş, 1976). Okullarda ölçme ve değerlendirme araçlarının çoğunluğu öğretmen tarafından hazırlanır; uygulama ve değerlendirme de öğretmen tarafından yapılır. Bu nedenle aracın ve değerlendirmenin niteliği de öğretmenin yeterliliğine bağlıdır.

Açık Uçlu Soruların Avantaj ve Dezavantajları

Açık uçlu ve kısa yanıtli sorulardan oluşan bir testin hazırlanması ve uygulanması kolaydır. Öğrenci sayısı az olduğunda kullanıma uygundur çünkü hazırlaması kolay olsa da puanlanma süreci uzundur. Çoktan seçmeli testlere göre; bu becerileri ölçecek şekilde hazırlandığı takdirde daha üst düzey bir zihinsel etkinlik ölçer çünkü sorunun yanıtını öğrencinin hatırlaması, bilgilerini kullanarak kurgulaması gerekir. Konu ile ilgili bilgilerini ölçmenin yanında öğrencilerin yazıları ve imla kurallarına ne kadar uydukları gözlemlenebilir. Yazılı yoklamalarda açık uçlu soruların sorulması zaman açısından kısıt yaratır bu nedenle de soru sayısı azdır; bu durumun güvenilirliği düşürmesi söz konusu olabilir. Soru sayısının azlığı nedeniyle özellikle kapsam geçerliği düşük olabilmektedir ancak ölçtüğü bilişsel düzey açısından (üst düzey becerileri de ölçtüğü takdirde) geçerlik yükselebilir (Karakaya vd., 2022). Yazılı yoklamalarda açık uçlu sorular kullanıldığı zaman öğretmen tarafından puanlamada yanlışlık oluşabilir. Bunun önüne geçilmesi için ayrıntılı puanlama anahtarı oluşturulması gerekir.

Kısa yanıtli sorularda da maddelerin hazırlanması kolaydır. Yanıtlamak kısa sürdüğü için kısa zamanda daha çok soru sorularak uzun yanıtli açık uçlu sorulara kıyasla güvenilirlik artmış olur. Yanıtlar tek sayıdan ya da bir kaç kelimeden oluştuğu için puanlama hatası az olur (Tekin, 1996).

Açık uçlu maddeler puanlama güvenilirliği açısından çoktan seçmeli testlere yaklaşmasa da hazırlanmasının ve uygulanmasının kolaylığı yanında üst düzey bilişsel alanlarda ölçüm yapılabilmesi ile özellikle sınıf içi ölçmelerde kullanılmaktadır. Günümüzde teknolojinin gelişimi ve YZ araçlarının kullanımının artması ile puanlamanın zorluğu ve yanlı olması gibi dezavantajları da ortadan kalkabilecektir (Tekin, 1996; Karakaya vd., 2022).

Puanlama Yöntemleri

Puan, ölçülmek istenen özelliğin bir parçasını yoklayan bir maddeye, o özelliğin bulunuş derecesine uygun olarak biçilen değer; örtük özelliğin öngörülen bulunuşluk derecesidir; puan bir anlamda bir ağırlıktır (Erkuş, 2019).

Puanlama ise, kullanılan ölçek türüne göre belirlenmesi gereken bir işlemdir. Ölçülmek istenen özelliği ölçebilmek için doğru şekilde hazırlanmış geçerli bir ölçme aracının puanlaması da aracın içeriğine, maddelerin yapısına ve test sonucunda verilecek geri bildirimine göre değişiklik gösterecektir. Seçme gerektiren (doğru yanlış, eşleştirme, çoktan seçmeli) testler ve açık uçlu (kısa yanıtlı, uzun yanıtlı) testler farklı şekillerde puanlandırılır. Bir testin hazırlanması, maddelerinin oluşturulması kadar puanlama yönteminin seçilmesi ve gerekli puanlama aracının hazırlanması da önemlidir.

Yanıtı yapılandırılmış test maddelerine verilen yanıtların önceden belirlenmiş kriterlere göre puanlanması işlemine madde puanlama denir. İkili ya da çoklu puanlama şeklinde yapılabilir. İkili puanlama, aslında sadece varlığı - yokluğu yokladığı için “gerçek süreksiz” yapıdaki özelliklerin ölçümünde kullanılmalıdır. Bir özelliğin varlığının veya yokluğunun araştırıldığı gözlem formlarında veya kontrol listelerinde kullanılması daha uygundur. Ancak çoktan seçmeli başarı testlerinde, başarı sürekli bir değişken olduğu halde 1 - 0 ikili (yapay süreksiz) puanlama yapılmaktadır (Erkuş, 2019). Çoktan seçmeli maddelerde, seçenekler arasında tek doğru yanıt vardır ve o yanıt dışında hangi seçeneğin işaretlendiği puan açısından önemli değildir. Kişi soruya doğru yanıt veriyse 1, yanlış seçenekleri işaretledi ya da soruyu yanıtlamadıysa 0 olarak puanlama yapılır. Puanlama yapılırken elle puanlama ya da dijital puanlama kullanılabilir. Nesnelliğin üst düzeyde olduğu bir puanlama yöntemi olduğu için elle yapılan puanlamalarla dijital puanlama arasındaki fark yalnızca puanlamanın kolaylığı olmaktadır (Downing & Haladyna, 2006).

Çoklu puanlama, bireyde bir yoğunluk içeren sürekli değişkenler için uygundur ve aslında tüm psikolojik değişkenler de böyledir (Erkuş, 2019). Çoklu puanlanan sorular, geleneksel tek doğru yanıtı çoktan seçmeli soruların geliştirilmiş biçimidir. Bu soru türünde, öğrencilerin birden çok doğru seçeneği belirleyebilmeleri için tüm olasıkları dikkatli bir şekilde değerlendirmeleri beklenir. Soru yapısı, ana metinden ve birkaç seçenekten meydana gelir; bu seçeneklerin her biri potansiyel olarak doğru yanıt olma özelliği taşıyabilir.

Bir ürün ortaya çıkarma, bir kompozisyon yazma, resim yapma, sunum yapma, bir maket-proje gibi bir ürün ortaya çıkarma ve aynı şekilde, açık uçlu maddeler bir değerlendirici gerektirir. Bu gibi yapılandırılmamış veya yarı yapılandırılmış tepkilerde puanlama sorunu ortaya çıkar. Bu sorunlara; testi alan kişinin vereceği cevapların tahmin edilemeyecek kadar çok olması, verilen cevaplar içinde hangilerinin ne kadar doğru olduğuna karar verilmesi gerekliliği, puanlayıcının yanlılığı gibi durumlar örnek verilebilir. Bunların önüne geçmek için bazı puanlama araçları (yanıt anahtarı, dereceleme ölçekleri ya da rubrikler) kullanılmaktadır.

YZ ile Otomatik Puanlama Yöntemleri

Teknolojinin eğitim alanında kullanımı arttıkça, ölçme ve değerlendirme süreçleri dijitalleşmiş ve YZ destekli otomatik puanlama sistemleri önemli hale gelmiştir. Otomatik puanlama, öğrenci yanıtlarının bilgisayar algoritmaları tarafından değerlendirilmesi sürecidir ve geleneksel puanlama yöntemlerinin zaman, maliyet ve tutarlılık sınırlılıklarına çözüm sunmaktadır. Bu alanda geliştirilmiş olan başlıca yaklaşımlar; makine öğrenmesi (machine learning), derin öğrenme (deep learning), büyük dil modelleri (Large Language Models), ince ayarlama (fine-tuning) ve az örnekli öğrenmedir (few-shot learning).

Makine öğrenmesi, geleneksel algoritma yapısında var olan koşula göre aksiyon alma durumları için aksiyon parametrelerinin algoritmalar tarafından farklı değişkenler ile denenerek, elde edilen sonuçlara göre modelin verimliliğini arttırabilecek şekilde mimari yapısının güncellenmesine olanak sağlamaktadır. Bu teknoloji, özellikle eğitimde ölçme ve değerlendirme alanında devrim niteliğinde değişiklikler yaratmakta ve geleneksel değerlendirme yöntemlerinin sınırlarını aşmaktadır. Derin öğrenme büyük verileri işleyebilme kabiliyetine sahip algoritmalar, bu da özellikle açık uçlu soruların

değerlendirilmesinde karmaşık metin analizlerinin yapılabilmesini mümkün kılar (Özgür, 2021).

Büyük dil modelleri, makine öğrenmesinin bir alt dalı olan derin öğrenme teknolojisinin en gelişmiş formudur. Geleneksel makine öğrenmesi yöntemlerinden farkı, milyarlarca parametre içermesi, geniş ölçekli eğitim verisi kullanması ve çoklu görev kapasitesine sahip olmasıdır. Bu özellikler, LLM'lerin spesifik eğitim gerektirmeden çeşitli dil görevlerini yerine getirmesini sağlamaktadır. Bu modellerin ince ayarlama (fine-tuning) süreçleri, belirli eğitim alanlarına özgü değerlendirme kriterleri için optimize edilmelerini sağlarken, az örnekli öğrenme (few-shot learning) yeteneği ise minimum örnekle maksimum performans elde etme imkanı sunmaktadır (Özgür, 2021).

Fine-tuning yaklaşımı, önceden eğitilmiş büyük dil modellerinin spesifik puanlama görevleri için yeniden eğitilmesi esasına dayanmaktadır (Devlin vd., 2018). Bu yöntem, derin öğrenmeye göre daha az veri ile iyi performans elde edebilmesine rağmen, teknik uzmanlık, GPU kaynaklarına erişim ve uzun eğitim süreçleri gerektirmektedir (Rogers vd., 2020). Son yıllarda ortaya çıkan az örnekli öğrenme yaklaşımı ise, büyük dil modellerinin sahip olduğu geniş bilgi tabanından yararlanarak, minimum örnek ile görev öğrenimini mümkün kılmaktadır (Brown vd, 2020). Bu yaklaşımda, model önceden belirlenmiş birkaç örnek ile görevin nasıl gerçekleştirileceğini öğrenir ve yeni örneklerle bu öğrenme temelinde yanıt verir.

Az örnekli öğrenme yaklaşımı, eğitim alanında önemli avantajlar sunmaktadır. Bu yaklaşım, binlerce veri gerektiren diğer yöntemlerin aksine sadece birkaç örnek ile etkili sonuçlar elde edilmesini sağlamaktadır. Teknik uzmanlık ve özel donanım gerektirmemesi, farklı derslere hızla uyumlanabilmesi ve düşük maliyet ile uygulama açısından önemli avantajlar sağlamaktadır. Ayrıca puanlama gerekçelerinin alınabilmesi ve öğretmen kontrolünün mümkün olması, şeffaflık ve açıklanabilirlik açısından eğitim paydaşları için güven verici özellikler oluşturmaktadır. Bu avantajlar, özellikle gelişmekte olan ülkelerde eğitim teknolojilerinin yaygınlaştırılması için önemli fırsatlar sunmaktadır.

Rubrik Kullanımı, Önemi ve Türleri

Rubrik, bir değerlendirme etkinliği için önceden belirlenen kriterleri açıklayan puanlama çizelgesidir. Bir görevi bileşen parçalarına ayırır ve bu parçaların her biri için hangi performans düzeylerinin yeterli veya yetersiz olduğuna dair ayrıntılı bir tanımlama

sağlar. Rubrikler; araştırma makaleleri, kitap eleştirileri, münazaralar, laboratuvar raporları, portfolyolar, grup çalışması, sözlü sunumlar ... gibi çok çeşitli ödev ve görevlere not vermek için kullanılabilir. Dereceli puanlama anahtarı aynı şekilde birden fazla görev ya da test için geçerli ise “genel”, verilen bazı görevler ya da testlere özel değerlendirmeler içeriyorsa “göreve özel” anahtar olarak ayrılır (Stevens & Levi, 2005; Kutlu vd., 2017).

“Dereceli puanlama anahtarları, öğrenci çalışmalarına ilişkin olarak ölçütlerin, bu ölçütlere yönelik tanımlamaların ve başarı düzeylerinin bulunduğu bir puanlama aracıdır” (Kutlu vd., 2017). Örneğin 5 düzeyli bir dereceli puanlama anahtarında, 4 ve 5 yüksek, 3 orta, 2 ve 1 ise düşük başarı düzeyini ifade etmektedir. Bu sayısal puanlar daha sonra toplanarak bir toplam nota dönüştürülür. Sayısal ifadeler yerine, başarı düzeyini anlatabilecek; “iyi, yeterli, geliştirilmeli, mükemmel” gibi sözel ifadeler de kullanılabilir.

Dereceli puanlama anahtarı, analitik ve holistik (bütünsel) olarak iki farklı şekilde hazırlanabilir. Farklı kaynaklarda anahtar oluşturma ile ilgili olarak sıralanan özelliklere bakıldığında ortak bazı özellikler ortaya çıkmıştır. Bunlar kısaca şöyle sıralanabilir (Haladyna (1997), Moskal (2000) ve Goodrich (2000)'in önerilerinden: Kutlu vd., 2017):

1. Beklenen performansa göre amacı belirlemek
2. Beklenen performansa göre ölçütleri tanımlamak
3. Tanımlanan ölçütlerin üst - orta - alt nitelik düzeylerine karar vermek
4. Nasıl bir puanlama anahtarı (bütünsel, analitik) kullanılacağına karar vermek.
5. Öğrencilere, hangi ölçütlere göre değerlendirilecekleri hakkında bilgi vermek ve ölçütleri öğrencilere açıklamak.
6. Belirlenen ilkelere göre uygulamayı yapmak ve çalışmalarını puanlama anahtarına göre puanlamak.
7. Sonuçları öğrencilerle paylaşmak, öncelikle yapabildiklerini sonra varsa eksiklikleri görmelerini ve kendilerini geliştirmelerini sağlamak.

Ancak tüm bu aşamalardan geçmek zorunlu değildir, öncelikle yapılan ölçümün sonuçlarının ne amaç ile kullanılacağı, öğrenciden beklenen başarı düzeyi, süre kısıtları ve ekonomik etkenler dikkate alınmalıdır.

Bütünsel puanlama anahtarında, görevi yapan (testi alan) kişinin gösterdiği performansın bütününe bakılarak puan verilmektedir. Performanstaki küçük hataların göz ardı edilebileceği puanlama şeklidir. Bazı sınırlılıkları vardır; örneğin farklı becerilerde yetkin olan iki ayrı birey benzer notlar elde edebilmekte, değerlendirilen kişinin sahip

olduğu üstün ve eksik yanlarını tespit etmede yetersiz kalmakta, ayrıca güvenilirlik açısından sorunlar taşımaktadır (Kutlu vd., 2017).

Kompozisyon (essay) şeklindeki görevlerde, sözlü sunumlarda bütünsel anahtarlar kullanılabilir. Ölçüm sonuçları ile elde edilecek puanlar kişilerin seçilmesi, yerleştirilmesi veya sıralanması gibi genel bir karar vermek amacı ile kullanılacaksa bütünsel puanlama anahtarı kullanmak uygundur. Bütünsel puanlama anahtarı, sonuçla ilgilidir. Sonuca ulaşana kadar geçilen bireysel basamaklarla değil toplam puan ya da performansla ilgilenir. Bunun yanında ölçülen yapının alt boyutları yoksa, bütün olarak bir özelliğin puanlanması hedeflenmişse bütünsel anahtar kullanılır (Karakaya, 2022).

Analitik puanlama anahtarı, öğrencilerin karşılaştığı zorlukları belirlemek ya da kişinin bireysel gelişimini öğrenmek amacıyla yapılan değerlendirmelerde kullanılır. Analitik puanlama anahtarı ile her performans kriteri ayrı ayrı değerlendirilir. Ayrıca birden fazla alt boyutu olan yapıların ölçülmesinde analitik anahtar kullanılır. Açık uçlu test maddelerinin puanlanmasında da kullanılabilir puanlama araçlarıdır. Analitik puanlama anahtarları ölçülecek her temel özellik için o özelliğe yönelik bir kriter belirlenmesi ile oluşturulduğundan bütünsel puanlama anahtarına göre geçerlik ve güvenilirliği daha yüksek puanlama sonuçları vermektedir. Ancak bazı sınırlılıkları vardır, örneğin hazırlanması uzun sürer ve puanlaması da zaman alır. Uzman kişiler tarafından hazırlanıp puanlaması gereklidir (Kutlu vd., 2017).

Analitik puanlama anahtarlarında, ölçütler eşit ağırlıklara sahip olabileceği gibi bazı ölçütler diğerlerinden daha önemli görülüyorsa ağırlığı diğer ölçütlere göre farklılaştırılabilir. Bunlar görevin niteliğine, görevi alan kişinin düzeyine ya da testi uygulayan kişinin geliştirmek istediği becerilere göre değişebilir (Kutlu vd., 2017).

Öğretmen Puanlamalarında Tutarlılık ve Güvenilirlik

Puanlayıcıların, puanlama sürecinde genel olarak düştükleri üç tür hata vardır; kişisel yanlılık hatası, ilk izlenim hatası, mantıksal hata. Kişisel yanlılık hatası; aşırı olumluluk (öğrenci performansını gerçek durumdan daha yüksek puanlanma eğilimi), aşırı olumsuzluk (var olan durumdan daha düşük puanlanma eğilimi) ya da merkezi eğilim hatası (puanlama yaparken öğrenciye yüksek veya düşük puan vermekten kaçınma) olarak ortaya çıkabilir. İlk izlenim hatası, öğretmenin önyargılı davranması ile oluşan, öğrencinin dikkat çeken olumlu ya da olumsuz bir özelliğinin puanlamaya etki etmesi olarak tanımlanabilir. Mantıksal hata ise öğretmenin, öğrencinin bir özelliği

konusundaki görüşünün başka bir özelliğinden etkilenmesidir. Dereceli puanlama anahtarı kullanılırken bu hatalardan herhangi biri yapılırsa puanlamanın güvenilirliği ve geçerliği zarar görecektir. Bunun sonucunda eğitimle ilgili alınan kararlar etkilenecek ve hatalı puanlamalara dayalı olarak verilen geri bildirimler işlevini yerine getirmeyecektir (Kutlu vd., 2017).

Yapay Zekâ ve Eğitim

Bilgisayarların kavramsal temelleri 1830'larda Charles Babbage'ın delikli kart kullanılarak programlanabilen "Analitik Makine"sine; 1936 yılında Alan Turing tarafından geliştirilen karmaşık matematiksel hesapların belirli bir düzenek tarafından yapılmasını sağlayan "Turing Makinesi"ne kadar dayanmaktadır ("Vikipedi," t.y.). Programlanabilir, genel amaçlı elektronik ilk bilgisayar ENIAC (Electronic Numerical Integrator And Computer) II. Dünya Savaşı sırasında Amerika'da Pennsylvania Üniversitesi'nde geliştirilmiştir. 1946 yılında kamuoyuna tanıtılan 30 ton ağırlığındaki bu bilgisayar, savaş sırasında topçu atış tablolarını hesaplamak için tasarlanmıştır ("Encyclopædia Britannica," t.y.).

1960'larda, nüfus ile birlikte eğitime olan talep artışının daha fazla okul yapmak ve öğretmen yetiştirmek ile karşılanamayacağı, eğitimin kendisinin verimli hale getirilmesi gerektiği belirtilmiştir (Skinner, 1968). Eğitimde kaynaklara ek olarak kullanılan görsel-işitsel materyallerin öğrencileri pasif alıcılar haline getirme tehlikesi taşıdığı vurgulanmıştır. Skinner'ın öğretim makineleri, öğrencilerin öğrenme sürecinde aktif rol almasını sağlayan farklı bir sistemdir ve edimsel koşullanma prensipleri temelinde daha etkili öğretim makineleri tasarlanabileceğini savunmuştur.

Eğitim alanında bilgisayar kullanımı 1960'larda başlamıştır. PLATO (Programmed Logic for Automatic Teaching Operations), Illinois Üniversitesi'nde öğrencilerin bireysel öğrenimlerini desteklemek amacıyla geliştirilmiştir. "Bilgisayar Destekli Öğretim" kavramının öncüsü olan sistem aynı zamanda günümüzde yaygın olan birçok teknolojinin ilk veya daha erken bir örneğidir.

1977 yılında Amerika'da Apple II, Commodore PET ve TRS-80 gibi ilk kişisel bilgisayarlar (mikrobilgisayar) üretilmiştir. Bilgisayara ulaşmanın kolaylaşması ile okullarda bilgisayar kullanımı yaygınlaşmaya başlamıştır. Aynı dönemde "Teknoloji okuryazarlığı" kavramı ortaya çıkmıştır. 1980'li yıllarda, okullarda bilgisayar

laboratuvarları kurularak temel bazı programlar öğretilmeye başlanmış, bilgisayar dersleri müfredatlara girmiştir.

1990'larda World Wide Web'in (genel ağ) gelişmesi, insanların bilgisayar kullanma amaçlarını, bilgisayarların insan hayatındaki yerini ve bilgiye erişimi tamamen değiştirmiştir. Uzaktan eğitim kavramı ortaya çıkmış ve bu alana özel yazılımlar artmaya başlamıştır. Eğitimde bilgisayar kullanımını destekleyecek programlar yaygınlaşmıştır.

YZ düşüncesi ilk bilgisayara kadar dayandırılabilse de tam kavram olarak ilk defa 1956 yılında Dortmund konferansında John McCarthy tarafından dile getirilmiştir. Bilgisayarın gelişimi, bilgiyi saklama ve sakladığı bilgiler arasında bağlantı kurabilme becerisi kazandırılması ile YZ sürecini hızlandırmıştır. 1997 yılında IBM'in geliştirdiği, düşünen makine Deep Blue ile Gary Kasparov karşı karşıya gelene kadar kimse bir insanın bir makineye yenileceğini düşünmemiştir (Say, 2018).

2000'li yıllarda interaktif akıllı tahtaların sınıflara girmesi ile dijital öğrenme ortamları oluşmaya başlamıştır. Okullarda bilgisayar sayısı artırılarak her öğrenciye bir bilgisayar olmasına yönelik projeler geliştirilmiştir. Akıllı telefonlar henüz yaygınlaşmasa da mobil teknolojiler yavaş yavaş gelişmiştir. 2010'lu yıllarda akıllı telefonlar ve tabletler hayatın içine girdiği gibi eğitimde de kullanılmaya başlanmıştır. Bu dönemde YZ destekli öğrenme sistemleri ortaya çıkmıştır. Çevrimiçi kurslar, öğretici videolar, ders anlatımları popülerleşmiştir.

COVID-19 ile birlikte tüm dünyada uzaktan eğitim zorunlu hale gelmiş ve bu sayede çevrimiçi öğrenme araçlarının kullanımı artmıştır. YZ ve makine öğrenimi tabanlı eğitim sistemleri yaygınlaşmıştır.

Yapay Zekâ ile Açık Uçlu Soruların Değerlendirilmesi ve Geri Bildirim

Bilgisayarların test maddelerinin puanlanmasında kullanımı, 19. yy'da görme engellilere yardımcı olmak amacıyla geliştirilen makinelerin evrimi ile optik karakter tanıma (OMR) teknolojisinin geliştirilmesi sayesinde olmuştur. 1930'larda, IBM şirketi test puanlaması için optik işaret algılama sistemleri üzerinde deneyler yapmış ve soruların yanıtının, yanıtlayıcıya hazır olarak verilen bir formdaki yuvarlak boşlukların kurşun kalem ile doldurulması sayesinde yanıtların otomatik olarak puanlanması sağlanmıştır ("Vikipedi," t.y.). Bu teknoloji, doğru/yanlış ya da çoktan seçmeli maddeler gibi yanıtı yapılandırılmış testleri hızlı ve güvenilir şekilde değerlendirmeyi sağlamıştır.

Makine öğrenmesi algoritmaları ile açık uçlu maddelerin otomatik puanlanmasının da önü açılmıştır. 1960 ve 1970’lerde başlayan denemeler sınırlı metin değerlendirmeleri yaparken 1990’larda teknolojinin hızlanması ile geliştirilen Educational Testing Service (Eğitimsel Test Hizmetleri - ETS) (Williamson, Xi, & Breyer, 2012), Ellis Page’in geliştirdiği Project Essay Grade (Proje Kompozisyon Değerlendirme - PEG) (Page, 1966) ve Pearson tarafından desteklenen Intelligent Essay Assessor (Akıllı Kompozisyon Değerlendirici - IEA) (Landauer, Laham, & Foltz, 2003) gibi ilk modern sistemler kompozisyon ve içerik değerlendirme için kullanılmıştır. English Language Services (İngilizce Dil Hizmetleri - ELS) ise dil eğitimi sağlayan bir kurum olarak bu teknolojilerin uygulama alanlarından biri olmuştur (ELS Educational Services, n.d.). Farklı türlerde algoritmalar sayesinde eğitilen makineler, verilen veri setleri ve öğretilmiş değerlendirme kriterleri sayesinde başarılı şekilde puanlama yapmıştır. YZ sistemlerinin alt kümesi olan bu algoritmaların denetimli öğrenme, denetimsiz öğrenme, yarı denetimli öğrenme, pekiştirmeli öğrenme, derin öğrenme, transfer öğrenimi, toplu öğrenme gibi çeşitli öğrenme şekilleri vardır (Alpaydın, 2020).

YZ, temel olarak insan zekasını taklit ederek insan gibi düşünebilmesi için geliştirilmiş; düşünebilen, öğrenebilen ve problem çözebilen sistemleri geliştirmeyi amaçlayan bilim dalıdır. Makine öğrenmesini de içine alan geniş bir alan olmasından dolayı 2000’lerde gelişen algoritmalar sayesinde hızla hayatımıza girmiştir. Makine öğrenmesinde genel yaklaşım algoritmanın veri kalıplarını tanıması ve bunlardan öğrenmesi iken YZ insanların düşünme sürecini taklit etmek üzerine kurulmuştur (Alpaydın, 2020).

YZ ile değerlendirilmiş maddelerden elde edilen sonuçlar, öğretmen tarafından puanlanmış olan testler gibi öğrenciye, veliye ve okula geri bildirim verilmesi amacı ile kullanılacaktır. Eğitimde geri bildirim; kazanılması beklenen davranış hakkında öğrenciye bilgi vermeyi, öğrenciyi yönlendirmeyi, güdülemeyi ve doğru davranışı pekiştirmeyi amaçlamaktadır. Öğrenci hakkında bilgi elde etmenin yanında eğitim politikalarının eksiklerini görmek için de geri bildirim önemlidir. Eğitim süreci ile ilgili verilen geri bildirimler eğitim sürecini, sınıf içi davranışları, öğrenme araçlarının ve hedef davranış için verilen sürenin yeterliliğini, hedeflenen davranışın öğrenci ve sınıf düzeyine uygunluğunu içermelidir (Sönmez, 2011). Bu durumda öğretmen tarafından puanlanan testlerin sonucunda verilen geri bildirimler bu kapsamı sağlarken YZ’nın vereceği geri bildirimlerin yalnızca o teste özgü olarak kalması öğrenciler ve eğitim politikalarını gözden geçirecek olan yöneticiler için yeterli olmayabilir.

Öğretmen ve Yapay Zekâ Değerlendirmelerinin Karşılaştırılması

Başarı testlerinin puanlanmasında dereceli puanlama anahtarı kullanmak, açık uçlu maddelerde hangi cevaba ne kadar puan verilmesi gerektiğini standartlaştırma açısından önemlidir. Bir maddenin yanıtlarını farklı puanlayıcıların aynı şekilde puanlayabilmesini sağlamak için dereceli puanlama anahtarı hazırlanmalıdır (Karakaya vd., 2022). Puanlama anahtarı ile yapılan değerlendirmede YZ ile öğretmen puanları arasında farklılar görülmesi başarı dışındaki etkenlerin öğretmenin verdiği notları etkilemesinin bir sonucu olarak görülebilir. Önemli olanın not değil öğrenme olduğunu savunan öğretmenlerin puanladıkları testlerde başarı dışındaki faktörleri göz önünde bulundurdıkları düşünülebilir ancak not vermede başarı dışındaki etkenler dikkate alınmamalıdır (Tekin, 1996). Bu durumda, öğrencilerin yazılı kağıtlarını önyargısız ve puanlama anahtarına göre tarafsız olarak değerlendiren YZ sistemlerinin daha gerçekçi puanlar vermesi olasıdır.

Güvenirlilik ve Geçerlik Karşılaştırmaları

Bir ölçme aracının işlevsel olması için ölçüm sonucunun gerçekleri yansıtması gereklidir. Gerçeklerden uzaklaşmaya neden olan ise ölçme sonuçlarına karışan hatalardır. Sabit hata (her ölçme sonucunda aynı etkiyi gösteren), sistematik hata (her ölçme sonucuna aynı şekilde etki etmeyen) ve tesadüfi hata (ölçme sonuçlarına rastgele karışan) olmak üzere 3 hata türü vardır. Sosyal bilimlerde gerçek sonuçlara ulaşmaktaki en önemli engel tesadüfi hatalardır. Ölçme aracında olması beklenen nitelikler bu hatadan arınık olmasıdır, yani geçerli ve güvenilir bir ölçme aracı olmasıdır (Tekin, 1996).

Geçerlik, amaca uygunluktur; “ölçme aracının, ölçmek istediği özelliği başka değişkenlerle karıştırmadan ölçebilme derecesidir” (Atılğan, 2018). Literatürde en sık kullanılan üç geçerlik türü vardır; ölçüt dayanaklı geçerlik, kapsam geçerliği, yapı geçerliği. Güvenirlilik ise bir ölçme aracının, ölçülmek istenen özelliği hep aynı şekilde ölçmesi, kararlı ve tutarlı olmasıdır. Klasik test kurminde sosyal bilimler için güvenirlilik tesadüfi hatalardan arınlık olarak tanımlanır (Büyüköztürk vd., 2021). Ölçme aracının farklı zamanlarda uygulandığında aynı puanı vermesi (kararlılık); maddelerin birbirleri ile ve aracın bütünü ile uyumlu olması (iç tutarlılık); madde sayısının yeterli olması (duyarlılık) ve çoktan seçmeli olmayan bir ölçme aracının, farklı puanlayıcılar tarafından

puanlandığında yakın sonuçlar vermesi (objektiflik) güvenilirlik açısından kontrol edilen özelliklerdir. Her bir özelliğin farklı hesaplama yöntemleri vardır.

Açık uçlu maddelerden oluşan bir ölçme aracının, farklı puanlayıcılar tarafından puanlandığında aynı ya da en azından yakın sonuçlar vermesi gereklidir. Bu güvenilirliği arttırmak için en önemli basamak ayrıntılı yanıt anahtarı hazırlamaktır. Puanlayıcıların verdiği puanları karşılaştırmak için değişkenlerin arasında ilişki olup olmadığına bakılır. Değişkenlerin arasında ilişki olması ya da olmaması durumuna korelasyon denmektedir. Bu hesaplama ile ilişki varsa ilişkinin yönü ve gücü belirlenir. Basit korelasyon iki değişken arasındaki ilişkiyi inceler, nedensellik belirtmez. Sonuçların tutarlı olup olmadığını anlamak için puanlayıcıların karşılaştırılmasında kullanılan farklı korelasyon hesaplamaları vardır. (Alıcı, 2014; Atılgan, 2018; Büyüköztürk vd., 2017; Crocker, & Algina, 1986; Doğan, 2022; Tekin, 1996)

Spearman Brown sıra farkları korelasyonu, Spearman's rho (ρ) olarak da adlandırılır. Serilerin sıra sayıları verilmişse ve sıraları önemliyse; değişkenlerden biri sıralı diğeri eşit aralıklı ya da oranlı ölçek düzeyinde ise ve en az eşit aralık düzeyinde olan ancak normallik varsayımının karşılanmadığı değişkenler (nonparametrik) var ise kullanılır. n toplam gözlem sayısı olan bir veri setinde x ve y değişkenlerinin sıralamaları üzerinden korelasyon hesabı yaparak iki puanlayıcının tutarlılığını ölçmeyi sağlar (Fielden ve Gibbons, 1993).

Kendall'in Konkordans katsayısı, Kendall's tau (τ) olarak adlandırılır. Spearman's ρ gibi sıralama ölçeği düzeyindeki iki farklı veri setinin arasındaki bağlantıyı incelemek için kullanılır; uyumsuz çiftlere göre uyumlu çiftlerin oranını verir. Küçük örneklerde daha iyi evren kestirimi yaptığı söylenebilir. Tau-a (iki nicelik arasındaki sıralı ilişkiyi ölçmek için), Tau-b (satır ve sütun sayısı eşit kare tablolar için) ve Tau- c (satır ve sütun sayısı eşit olmayan dikdörtgensel tablolar) şeklinde farklı hesaplamaları vardır. İki veri setinde sıralaması aynı olan verilerin olasılığı ile sıralaması farklı verilerin olasılığı arasındaki farkı gösterir (Fielden ve Gibbons, 1993).

Cohen'in (ağırlıklandırılmamış) kappa katsayısı ve ağırlıklandırılmış kappa katsayısı, farklı puanlayıcıların aynı test için verdiği puanların veya tek bir puanlayıcının farklı zamanlarda aynı puanlama kriterlerini kullanarak verdiği puanların uyumlarını belirlemek amacı ile kullanılır. Bu çalışmada YZ ve öğretmen puanlamalarının karşılaştırılmasında ağırlıklandırılmış kappa (QWK) istatistiği kullanılmıştır. İzleyen bölümde ilgili araştırmalar sunulmuştur.

İlgili Araştırmalar

YZ gelişimi, eğitimde YZ kullanımı, test sonuçlarının YZ ile değerlendirilmesi ve öğrencilere YZ destekli sistemlerin verdiği geri bildirimler üzerine alanda yapılmış ulusal ve uluslararası birçok çalışma bulunmaktadır. Bu çalışmalar önce ulusal ardından uluslararası olacak şekilde kronolojik sırada sunulmuştur.

Ulusal Literatürdeki Çalışmalar

Türkiye'de eğitimde YZ konusunda yapılan araştırmalar, özellikle 2020 yılından itibaren hız kazanmış ve son yıllarda bu alandaki çalışmalarda bir artış olduğu görülmüştür (Akdeniz & Özdiç, 2021). Yapılan çalışmalarda, araştırmacılar YZ teknolojilerinin eğitim alanındaki potansiyelini farklı boyutlarıyla ele almışlardır. Bu bağlamda, ulusal literatürde yer alan başlıca çalışmalar ve yaklaşımları aşağıda incelenmektedir.

Arf (1959), makinelerin düşünme kabiliyetleri ve bu süreçteki teknik mekanizmaları ele almıştır. İnsanların farklı etkilere farklı tepkiler vermesi gibi, makinelerin de programlanmış komutlara tepki verdiğini ifade etmiş ve bu tepkileri, makinelerin düşündüğünü gösteren belirtiler olarak kabul etmiştir. Ancak, Arf makinelerle yönelik bir eleştiride bulunmuştur: “Makinelerin çözme yeteneği, makine yapılırken programlanan problemlerle sınırlıdır. Bu sebeple, makineler insan beyninin anlama yeteneğine sahip değildir, yani yeni durumlara uyum sağlayabilme yeteneği yoktur.” Arf, insan beyninin daha önce karşılaşmadığı problemleri çözebilme yeteneğine sahip olması gibi makineler de yapılırken düşünülmemiş olan problemleri çözebilecek midir ve bu nasıl mümkün olabilecektir diye sormuştur.

Aradan geçen altmış beş yılda yaşanan teknolojik gelişmeler ile artık günümüzde düşünebilen ve problem çözebilen makineler vardır. Bu teknolojiler birçok alanda insanların işini kolaylaştırmakta ve yanında bazı sorular getirmektedir. Yapay zekanın kullanıldığı alanlarda insanların işlerini ellerinden alacağı korkusu, bağımsız düşündükleri zaman insanlarla karşı karşıya gelebilecekleri endişesi gibi distopik düşüncelerin yanında yapay zeka ile hayatın kolaylaşması ve gelişmenin hızlanması gibi olumlu beklentiler daha gerçekçi görünmektedir. Eğitimde yapay zeka uygulamaları, teknolojinin diğer alanlarda olduğu gibi, çağımızın önemli araştırma konuları arasında

yer almaktadır. Bu alandaki akademik çalışmalar özellikle 2000'li yıllardan itibaren hız kazanmıştır.

Kazu ve Özdemir'in (2009) çalışmasında öğrencilerin bireysel farklılıklarını dikkate alarak YZ teknolojilerinin eğitimde nasıl kullanılabileceği araştırılmaktadır. Çalışma, günümüzde hızla gelişen bilim ve teknolojinin, sürekli değişen ekonomik, sosyal ve kültürel durumlarla birlikte bireylerin yetenekleri ve bireysel farklılıklarına göre eğitim almalarının gerekliliğini vurgulamaktadır. Çalışma, eğitimde her bireyin farklı öğrenme özelliklerine sahip olduğunu ve bu nedenle her öğrenciye aynı eğitimin verilmesinin bireysel farklılıkları göz ardı ettiğini belirtir. Bireylerin farklılıklarını anlamak için yapılan çalışmalarda, özellikle Bilişsel Giriş Davranışları ve Duyuşsal Giriş Özellikleri gibi özelliklerin göz önünde bulundurulması gerektiği vurgulanmıştır. Ayrıca, zeka türleri ve öğrenme stilleri gibi bireysel özelliklerin dikkate alınması, öğretim tasarımı çalışmalarında önemli bir yer tutmaktadır. Ancak bu özelliklerin tespitinde kullanılan ölçeklerin değerlendirilmesi, eğitimciler tarafından yapılmaktadır. Yapay zeka teknolojilerinin, bu ölçeklerin değerlendirilmesinde kullanılma potansiyeli bulunmaktadır. Çalışma, öğrencilerin bireysel özelliklerinin bilgisayar sistemleri tarafından belirlenmesinin, zaman ve finansman açısından daha verimli olacağını belirtmektedir. Bireysel özellikleri belirleyecek personelin farklı zamanlarda ve durumlarda farklı yorumlar yapabileceği ve objektif değerlendirmeden uzaklaşabileceği düşünüldüğünde, bilgisayar sistemlerinin bu noktada daha güvenilir olduğu ifade edilmektedir. Yapay zeka teknolojilerinin kullanılmasıyla, öğrencilerin bireysel özelliklerinin daha doğru ve objektif bir şekilde belirlenebileceği, bu sayede eğitimde bireysel farklılıkların daha etkin bir şekilde dikkate alınabileceği sonucuna varılmıştır (Kazu & Özdemir, 2009).

Endüstri 4.0 sürecinin eğitim ve gelişim faaliyetleri üzerindeki etkilerini, özellikle yeni nesil bilişim teknolojilerinin bu sürece entegrasyonunu inceleyen bir başka çalışma, teknolojinin hızlı gelişimine ayak uydurmanın ve köklü dönüşümlere uyum sağlamanın ancak güncel eğitim ve geliştirme uygulamaları ile mümkün olabileceğini vurgulamaktadır (Demirkaya ve Sarpel, 2018). Çalışma, Endüstri 4.0'ın bilişim ve iletişim teknolojilerindeki gelişmeler, veri toplama, paylaşma, otomasyon ve üretim teknolojilerindeki yeni yaklaşımları kapsadığını belirtmekle birlikte bu sürecin işsizlik oranlarını artırabileceği ve insanların işlerini kaybetmelerine neden olabileceği üzerine endişeler de dile getirilmiştir.

Bulut bilişim ile ilgili olarak en önemli konunun bilgiye her yerden erişimi kolaylaştıran ve işbirliği için yeni imkanlar sunan bir teknoloji olması belirtilmiştir. Bulut bilişim teknolojisinin, bilgiye zamandan ve mekandan bağımsız olarak erişimi kolaylaştırdığı, maliyetleri düşürdüğü ve grup çalışmalarını desteklediği vurgulanmıştır. Ancak güvenlik sorunları ve endişeler de bu teknolojinin kullanımıyla ilişkilendirilmiştir. Yapay zeka teknolojisinin ise öğrenme ve öğretme süreçlerinde özelleştirilmiş eğitim sunarak büyük bir potansiyel taşıdığı ifade edilmiştir. Akıllı Eğitim Sistemleri gibi yapay zeka destekli teknolojilerin, eğitimde kişiselleştirilmiş yaklaşımları mümkün kıldığı belirtilmiştir. Bu çalışma, eğitim ve gelişim faaliyetlerinde yeni nesil bilişim teknolojilerinin rolünü vurgulayarak, bu teknolojilerin insan kaynakları yönetimi, öğrenme ve öğretme süreçleri üzerindeki etkisini detaylı bir şekilde incelemiştir. Bu teknolojilerin, eğitim ve geliştirme süreçlerine entegre edilmesinin, öğrenme ve öğretme faaliyetlerinin kalitesini artırabileceği, ancak aynı zamanda bazı zorluklar ve endişeleri de beraberinde getirebileceği vurgulanmıştır (Demirkaya ve Sarpel, 2018).

Arslan (2020), makalesinde eğitim alanında YZ'nin etkisini ve rolünü tartışmaktadır. YZ, bu yüzyılın teknolojide itici gücü olarak kabul edilmektedir. Öğrenme, akıllı tahminlerde bulunma, karmaşık problemleri çözme ve değişken koşullara uyum sağlama gibi özellikleriyle, YZ, eğitimde bilginin yönetiminden öğretim sürecine kadar geniş bir yelpazede yer almaktadır. Eğitimde YZ, veri tabanlı uygulamaların yanı sıra kişiselleştirilmiş ve diyalog tabanlı eğitim sistemleri, keşifçi eğitim, eğitimde veri madenciliği, öğrenci makale analizi, akıllı ajanlar, chatbotlar, özel ihtiyaçlı çocuklar için eğitim gibi çok çeşitli alanlarda yer almaktadır. Ayrıca, okulların ve üniversitelerin yönetsel yönünü destekleyerek, ders programları, personel yönetimi, sınav yönetimi gibi alanlarda da katkıda bulunmaktadır.

YZ, eğitimde çeşitli uygulamalar sunmaktadır. Özellikle uzman sistemler, belli bir alanda uzman kişilerin görevlerini yapay zeka algoritmaları kullanarak yerine getiren bilgisayar programlarıdır. Uzaktan eğitimde öğrenciye özelleştirilmiş geri bildirim ve problem çözme yaklaşımları sunar. Bu makaleye göre eğitimde yapay zekanın ilk uygulamalarından biri, 1920'de Sidney L. Pressey tarafından gerçekleştirilmiştir. Pressey, çoktan seçmeli testlerin, öğrencilerin başarılarını değerlendirmenin yanı sıra öğrenmelerini de pekiştirmek için kullanılabileceğini öne sürmüştür. YZ, eğitimde uzman sistemler gibi çeşitli uygulamaları içermektedir. Uzman sistemler, belli bir alanda uzman kişilerin yerine YZ algoritmaları kullanarak görevleri tamamlayan bilgisayar programlarıdır. Bu sistemler, bilgi tabanlı, çıkarım/karar mekanizması ve arayüz gibi

temel modüllere ihtiyaç duyarlar ve etkin bir şekilde kullanılmaları için bu modüller gereklidir. YZ, öğrenci deneyimini geliştirmenin yanı sıra eğitim kurumlarının yönetim süreçlerini kolaylaştıran yenilikçi çözümler ve kişiselleştirilmiş öğrenme yaklaşımları sunmaktadır. Bu, eğitim içeriğinin sunulma ve yönetilme biçiminde önemli bir değişikliğe işaret etmektedir (Arslan, 2020).

Akdeniz ve Özdiç (2021), YZ'nin eğitimde kullanımını ve bu konuda Türkiye'de yapılan çalışmaları sistemli bir şekilde incelemeyi amaçlamaktadır. Araştırmada, Ulusal Tez Merkezi ve Türkiye adresli dergiler taranarak elde edilen 37 çalışma, betimsel içerik analizi yöntemi ile incelenmiştir. Bulgular, son yıllarda eğitimde yapay zeka konusunda yapılan çalışmaların sayısında hızlı bir artış olduğunu ve özellikle üniversite öğrencileri ile çalışmaların yapıldığını göstermektedir. Yapay zeka uygulamalarından en yaygın kullanılanlar arasında zeki öğretim sistemleri, uyarlanabilir öğrenme sistemleri ve öneri sistemleri yer almaktadır.

Çalışmanın bulgularına göre, 1999-2018 yılları arasında yapılan çalışmaların çoğunluğu tasarım ve geliştirme yöntemi ile gerçekleştirilmiştir, bu durum yapay zeka alanında yeni teknolojilerin geliştirildiğini göstermektedir. Yüksek lisans ve doktora tezlerinin en çok yayın türü olduğu belirlenmiş, bunu eğitim dergileri ve eğitim fakülteleri dergileri takip etmektedir. Yapay zeka tekniklerinin en çok ele alındığı alan zeki öğretim sistemleri olarak karşımıza çıkmaktadır. Bu sistemler, yapay zeka tekniklerini kullanarak öğrencinin öğrenme sürecindeki performansını değerlendirir ve eğitim programını buna göre düzenleyebilir (Akdeniz ve Özdiç, 2021).

Çalışmalarda ele alınan çalışma gruplarına göre yapılan sınıflandırmada, üniversite öğrencilerinin çoğunluğu oluşturduğu ve tüm yaş gruplarına yönelik çalışmaların da yapıldığı tespit edilmiştir. Araştırmada, yapay zeka ile ilgili çalışmaların yayın türlerine göre dağılımı incelenmiş ve tez çalışmalarının çoğunlukta olduğu belirlenmiştir. Bu bulgular, yapay zekanın eğitimde kullanılmasının Türkiye'de son yıllarda gelişen bir araştırma alanı olduğunu göstermektedir. Çalışma, yapay zekanın eğitimde kullanımına ilişkin genel eğilimleri ve eksiklikleri ortaya koymakta ve bu alandaki araştırmalar için rehberlik etmeyi amaçlamaktadır (Akdeniz ve Özdiç, 2021).

Çetin ve Aktaş (2021)'in makalesinde, YZ'nin eğitim alanında nasıl bir etki yaratacağı ve bu alanda nasıl kullanılabileceği üzerine derinlemesine bir inceleme sunulmaktadır. Makale, YZ'nin eğitim alanındaki potansiyelini, olası faydalarını ve endişelerini ortaya koyarken, aynı zamanda bu teknolojinin mevcut eğitim sistemine entegrasyonunu ve gelecekteki olası senaryoları ele almıştır. YZ, eğitimde henüz tam

anlamıyla entegre olmamış olsa da, bu alandaki hızlı gelişimlerin yakın gelecekte eğitimin temel paydaşlarını etkileyeceği öngörülmüştür. Eğitim örgütleri, öğretmenler, öğrenciler ve veliler açısından yapay zekanın hem avantajları hem de dezavantajları detaylı bir şekilde incelenmiştir. Özellikle öğrenci ve öğrenme odaklı gelişmelerin öne çıktığı, kişiselleştirilmiş öğretim ve öğretim asistanlığı gibi konularda yapay zekanın önemli fırsatlar sunduğu belirtilmiştir. Makale, YZ'nin eğitim alanındaki geleceğini iki ana senaryo üzerinden değerlendirmiştir: “Sınıf Yönetiminde Yapay Zeka” ve “Okul Yönetiminde Yapay Zeka”. Bu senaryolara göre; yapay zeka öğretmenin ve okul müdürlerinin yerini alabilecek ve öğrencilerle ilişkiler, ölçme-değerlendirme, müfredat ve ders planlaması gibi konularda önemli faydalar sağlayabilecek, ancak aynı zamanda bazı endişeleri de beraberinde getirecektir. Özellikle yapay zekanın insan öğretmenlerin pedagojik ve duygusal yaklaşımlarını tam olarak taklit edemeyeceği ve bu nedenle öğretmenin yerini tam anlamıyla alamayacağı, ancak yardımcı bir rolde faydalı olabileceği vurgulanmaktadır. Okul güvenliği, akademik başarı, mali yönetim gibi konularda önemli katkılar sunabilecek ancak yine de, yapay zekanın insan liderlerle etkileşimi anlama ve grup üyelerini bütünün bir parçası gibi hissettirme konusunda bazı eksiklikler yaşayacağı düşünülmektedir.

Yapay zeka, idari koordinasyon ve kontrol gibi işleri otomatikleştirebilirken, insani ilişkiler ve karar verme gibi konularda insan liderlerin yerini tutamayacak gibi görünmektedir. Her iki senaryoda da, yapay zekanın eğitim alanında önemli faydalar sağlayabileceği, ancak insan öğretmenlerin ve liderlerin sahip olduğu benzersiz yeteneklerin yerini alamayacağı vurgulanmaktadır. Bu nedenle, yapay zekanın eğitimde yardımcı bir rol oynayabileceği, ancak öğretmen ve liderlerin yerini tam anlamıyla alamayacağı sonucuna varılmıştır (Çetin ve Aktaş, 2021).

İşler ve Kılıç (2021), eğitim alanında YZ teknolojilerinin kullanımı ve etkileri üzerine odaklanmaktadır. Makalede, YZ'nin günümüzde hayatımızın vazgeçilmez bir parçası haline geldiği ve birçok sektörde başarılı bir şekilde kullanıldığı belirtilmiştir. Eğitim sektöründe bu teknolojilerin kullanılmaya başlandığı ve çalışmaların, gelecekteki araştırmalar için kaynak niteliğinde olabileceği ifade edilmiştir. YZ teknolojilerinin eğitimdeki potansiyelini ve bu alanda sağlayabileceği faydaları ayrıntılı bir şekilde ele almaktadır. YZ uygulamalarının, öğrencilerin bireysel öğrenme ihtiyaçlarına göre kişiselleştirilmiş öğretim imkânları sunarak, öğrencilerin öğrenme stilleri, yetenekleri ve ilerlemeleri hakkında yararlı bilgiler sağlayabileceği ve öğretim yöntemlerini öğrencilerin bireysel ihtiyaçlarına göre nasıl özelleştirebileceği konusunda önerilerde bulunabileceği

belirtilmiştir. Ayrıca, YZ uygulamalarının öğrencilerin kendi düzey ve öğrenme hızına göre öğrenme imkânı sunması, bu uygulamaların en büyük artısı olarak görülmektedir.

Makalede, YZ uygulamalarının, eğitimde 21. yüzyıl becerilerinin kazandırılmasını kolaylaştıracağı ve bu becerilere sahip öğrencilerin gelecekte ülkelerin kalkınması adına önemli bir role sahip olacakları ifade edilmektedir. YZ uygulamalarının eğitime getirdiği faydalar göz önünde bulundurulduğunda, okulların tamamen yapay zekâ destekli eğitime geçmeleri önerilmektedir (İşler ve Kılıç, 2021).

Irmak (2023) doktora tezinde, YZ teknolojilerinin eğitimde, özellikle uzaktan eğitimde, nasıl kullanılabileceğini araştırmaktadır. Çalışma, COVID-19 pandemisi sonrası, dünya genelinde eğitimin büyük ölçüde uzaktan eğitime dönüştüğü bir dönemde gerçekleştirilmiştir. Bu dönemde ortaya çıkan büyük veri, YZ algoritmalarının öğrenci akademik performanslarını tahminlemedeki başarısını değerlendirmek için önemli bir fırsat sunmuştur. Tezin temel amacı, YZ yöntemlerini kullanarak uzaktan eğitimde karşılaşılan sorunları tespit etmek ve öğrencilerin akademik performanslarını tahmin etmektir.

Öğrencilerin akademik performanslarının tahmin edilmesi amacıyla, Destek Vektör Makineleri, Yapay Sinir Ağları, Lojistik Regresyon, Random Forest, k-En Yakın Komşu ve Naive (Basit) Bayes Sınıflandırıcı gibi çeşitli sınıflandırma algoritmaları kullanılmıştır. Çalışmada, bir devlet üniversitesindeki tüm programlarda yer alan 1000 farklı öğrenci ve toplamda 7716 dersi kapsayan bir öğrenci veri seti kullanılmıştır. Modellerin performansları tabakalı çapraz geçiş ve tabakalı rastgele örnekleme yöntemleri kullanılarak değerlendirilmiştir. Sonuç olarak, teknolojinin eğitimdeki önemi ve yapay zekanın potansiyeli ortaya konulmuştur. Eğitimde dijital dönüşüm ve yeni teknolojilerin kullanımı, öğrencilerin iletişim, kritik düşünme, yaratıcılık, işbirliği, dijital okuryazarlık gibi becerilerini geliştirme potansiyeline sahiptir. Endüstri 4.0'ın etkilediği ve dijital dönüşümün en çok yaşanması beklenen sektörlerden biri olan eğitim, nitelikli bilgi ve iletişim teknolojilerindeki değişimi yansıtan, gençlerin bilgi ve becerilerini artıran, yaratıcı ve yenilikçi yönlerini güçlendiren bir yaklaşımla ele alınmalıdır (Irmak, 2023).

Akan (2024) tezinde eğitim alanında ölçme ve değerlendirme süreçlerinde YZ teknolojilerinin rolünü inceleyerek bu araçların pedagojik değerini ve sınırlarını ortaya koymaktadır. İçerik analizi metodolojisiyle yürütülen bu araştırma, YZ temelli değerlendirme sistemlerinin öğretmenlerin rolünü ortadan kaldırmayı değil, onların mesleki uygulamalarını zenginleştirmeyi amaçladığını belirtmektedir. İncelenen

uygulamalarda YZ teknolojilerinin sağladığı temel faydalar arasında eş zamanlı dönüt mekanizmaları, kişiselleştirilmiş değerlendirme olanakları ve daha nesnel puanlama süreçleri öne çıkmaktadır. Araştırma, YZ uygulamalarının eğitimcilerin değerlendirme süreçlerindeki iş yoğunluğunu azaltma, süreçleri hızlandırma ve değerlendirmede tutarlılığı artırma potansiyelini vurgulamaktadır. Ancak çalışmada, bu teknolojilerin insan muhakemesinin tamamen yerini almasının etik ve eğitimsel açıdan sorunlu olabileceği belirtilmiş; dolayısıyla YZ araçlarının destekleyici ve tamamlayıcı sistemler olarak kullanılmasının daha uygun olacağı savunulmuştur. Ayrıca, öğrenme süreçlerindeki bireysel farklılıkları dikkate alabilen bu teknolojilerin, geleneksel değerlendirme yaklaşımlarının sınırlılıklarını aşmada eğitimcilere önemli katkılar sağlayabileceği ifade edilmiştir. Çalışmanın vardığı sonuç, YZ destekli değerlendirme teknolojilerinin eğitim pratiğinde öğretmenlerle işbirliği içinde çalışacak şekilde entegre edilmesi gerektiği yönündedir. Akan'ın araştırması, bu teknolojilerin insan uzmanlığının yerine geçecek bir alternatif olarak değil, onu güçlendirecek bir destek mekanizması olarak yapılandırılmasının önemini vurgulamaktadır.

Kara (2025), doktora tezinde, teknik uygulama ve kullanıcı deneyimlerini ele almaktadır. Çalışma, doğal dil işleme teknolojilerinin eğitim alanındaki pratik uygulamalarına odaklanmaktadır. Kara tarafından geliştirilen değerlendirme sistemi, LLM'leri temel alarak öğrenci yanıtlarını belirli kriterlere göre analiz etme becerisine sahiptir. Özellikle belirtilen nokta ise söz konusu sistemin, eğitimcilerin yerine geçmek için değil, onların mesleki çalışmalarını destekleyici bir unsur olarak tasarlanmış olmasıdır. Sistemin temel hedefleri arasında eğitimcilerin iş yükünü hafifletmek, dönüt sürelerini optimize etmek ve değerlendirme süreçlerinde standardizasyonu arttırmak bulunmaktadır. Araştırma kapsamında yapılan on iki kişilik görüşme grubundan elde edilen verilere göre; katılımcılar, YZ sistemlerinin insan kaynaklı değerlendirme hatalarını minimize edebileceğini belirtmekle birlikte, insan denetiminin mutlaka sürecin bir parçası olması gerektiğini vurgulamışlardır. Ayrıca, değerlendirme sisteminin karar mekanizmalarının daha anlaşılır hale getirilmesi ve değerlendirme kriterlerinin daha net biçimde ortaya konması yönünde öneriler sunulmuştur. Bu araştırma, YZ temelli değerlendirme teknolojilerinin eğitim alanında asıl karar verici olarak değil, eğitimcilerin profesyonel yargılarını güçlendiren tamamlayıcı mekanizmalar olarak konumlandırılmasının uygun olacağını göstermektedir. Çalışma, insan eğitimci ve YZ sistemlerinin birlikte çalışmasına dayalı bir pedagojik model önermektedir.

Ulusal alan yazın incelendiğinde, YZ'nin soruların değerlendirilmesinde kullanılmasının sınırlı olduğu; yapılan çalışmaların çoğunun öğrencilerin bireysel özelliklerinin belirlenmesi, zeki öğretim sistemleri ve uyarlanabilir öğrenme sistemleri, kişiselleştirilmiş eğitim uygulamaları, uzaktan eğitimde akademik performans tahmini ve genel YZ destekli eğitim uygulamaları konularına odaklandığı görülmektedir. Açık uçlu soruların değerlendirilmesinde YZ kullanımını inceleyen çalışmaların sayısının ise oldukça sınırlı olduğu belirlenmiştir. Bu alanda yalnızca Akan (2024) ve Kara (2025) tarafından yapılan çalışmalar, ölçme ve değerlendirme süreçlerinde YZ destekli uygulamaları ele almakta ve açık uçlu soruların otomatik değerlendirilmesi konusuna değinmektedir. Bu durum, açık uçlu matematik sorularının değerlendirilmesinde YZ ve öğretmen puanlamalarının karşılaştırılması konusunda ulusal literatürde önemli bir boşluk olduğunu göstermektedir.

Uluslararası Literatürdeki Çalışmalar

Uluslararası literatürde, YZ destekli değerlendirme sistemleri konusunda uzun bir geçmişe dayanan kapsamlı araştırmalar bulunmaktadır. Özellikle 2000'li yıllardan itibaren, doğal dil işleme teknolojilerindeki gelişmeler ve makine öğrenmesi algoritmalarındaki ilerlemeler, açık uçlu soruların otomatik değerlendirilmesi alanında önemli çalışmaların yapılmasını mümkün kılmıştır. Bu çalışmalar, hem yazılı yanıtların analiz edilmesi hem de öğretmen değerlendirmeleriyle karşılaştırmalı analizler açısından zengin bir bilgi birikimi oluşturmuştur. Uluslararası literatürde yer alan temel çalışmalar ve yaklaşımları aşağıda ele alınmaktadır.

Dumas vd. (2021), katılımcıların ıraksak düşünme görevlerine verdiği özgün yanıtlarını puanlamada metin madenciliği modellerinin kullanımına yönelik artan ilgiyi ele almaktadır. Hem insan puanlayıcılardan hem de metin madenciliği sistemlerinden elde edilen puanların güvenilirliği ve geçerliliği karşılaştırılmıştır. Çalışmada, metin madenciliği modellerinin (özellikle GloVe 840B sistemi) ıraksak düşünmeyi değerlendirmede güvenilir ve geçerli sonuçlar sağlayabileceği sonucuna varılmaktadır. Araştırmanın sonuçları, yanıtın katılımcılar tarafından oluşturulduğu açık uçlu maddelerden oluşan testlerde insan değerlendiriciler kadar metin madenciliği modellerinin de etkili bir şekilde kullanılabileceğini göstermektedir. Bu sonuç, özellikle büyük veri setlerinin analiz edilmesi gereken durumlarda, zaman ve maliyet açısından önemli avantajlar sunabilir.

Raz vd. (2024), soru sormanın öğrenme, yaratıcılık ve bilişsel gelişim için önemli olduğunu vurgulamaktadır. Yapılan farklı çalışmalarda açık uçlu sorulara verilecek yanıtların karmaşıklığının azaltılabilmesi için Bloom taksonomisi gibi çerçeveler kullanılmış ancak karmaşıklığın ölçümü zorlu olmaya devam etmiştir. Doğal dil işleme alanındaki son gelişmeler, psikolojik yapıların YZ destekli bir şekilde otomatik puanlamasını, özellikle yaratıcılık derecelerinin tahmin edilmesini mümkün kılmıştır. Ancak, açık uçlu soruların otomatik değerlendirilmesi üzerine şimdiye kadar çok az araştırma yapılmıştır. Bu makalede yazarlar, yine Bloom taksonomisine dayanarak açık uçlu soru karmaşıklığının insan derecelendirmelerini doğru bir şekilde tahmin eden LLM kullanımını araştırmışlardır. Oluşturulan sorulara insanların verdiği yanıtlar kullanılarak Bloom taksonomisine göre soruları puanlamak üzere bir LLM'yi eğitmişler ve bulgular, LLM tarafından verilen puanların insan puanlayıcıların derecelendirmeleriyle güçlü bir korelasyon ($r = .73$) göstermiştir. Bu çalışma, açık uçlu sorulara verilen yanıtların karmaşıklığının otomatik değerlendirilmesinde LLM'nin önemini vurgulamakta ve bu alandaki otomatik ve güvenilir ölçümleri teşvik etmektedir.

Martínez - Comesaña vd., (2023), YZ'nin eğitimde, özellikle ilkökul ve ortaokul seviyelerindeki öğrenci değerlendirmelerine olan etkisini incelemektedir. Makale, YZ'nin eğitimde çeşitli yönlerde entegrasyonunun, bilgisayar bilimi, istatistik, psikoloji ve eğitim gibi disiplinlerin birleşimiyle oluşturduğu çok disiplinli bir alanı ortaya çıkardığını belirtmiştir. Özellikle öğrenci değerlendirme süreçlerini nasıl geliştirebileceğini gösteren mevcut çalışmaları sentezlemeyi amaçlamaktadır. Analize dahil edilen dokuz orijinal çalışma, 2010-2023 yılları arasında yayımlanmış ve YZ'nin bu alt eğitim seviyelerindeki öğrencilerin performansını tahmin etme, değerlendirmeleri daha objektif hale getirme, eğitim robotlarını kullanma ve dersleri daha çekici kılan özel faktörleri belirleme gibi başlıca katkıları vurgulanmıştır. YZ'nin, eğitimde öğrenci değerlendirmesini geliştirmek, öğretmenlerin görevlerini daha etkili ve verimli bir şekilde yerine getirmelerini kolaylaştırmak ve kişiselleştirilmiş öğrenme, adaptif değerlendirmeler, akıllı öğretim sistemleri, otomatik derecelendirme, sanal gerçeklik ve artırılmış gerçeklik gibi çeşitli eğitim unsurlarını geliştirebileceği belirtilmiştir.

YZ'nin eğitimde kullanımının, yönetim, öğretim ve öğrenme gibi farklı alanlarda önemli etkileri olduğu ve potansiyel öğrenme eksikliklerini keşfetmeyi ve bunlara erken müdahale etmeyi, öğrencinin ihtiyaçlarına uygun müdahaleler yapmayı ve verileri inceleyerek kariyer yollarını tahmin etmeyi mümkün kıldığı vurgulanmıştır. Ancak, yapay zeka destekli eğitimin sunduğu tüm olasılıklara rağmen, kullanımının etkin bir

şekilde uygulanabilmesi için araştırma ve çalışma yapmaya devam etmenin önemli olduğu ifade edilmiştir. YZ'nin farklı tipleri ve eğitimdeki uygulamaları açıklanmıştır (analitik, fonksiyonel, etkileşimli, metinsel ve görsel yapay zeka). Bu tiplerin her biri gerçek problemlere çözümler sunma yeteneğine sahiptir. Eğitimde yapay zeka alanı, bilgisayar bilimi, istatistik ve eğitim gibi bilgi dallarını kapsar ve bu alandaki disiplinlerarasılık, bilişsel psikoloji ve sinirbilimden gelen katkılarla zenginleştirilir. Makalenin temel sonucu, yapay zekanın karmaşıklığına rağmen, sistematik araştırmanın, özellikle ilkökul veya ortaokul gibi alt seviyelerdeki öğrenci değerlendirmelerini geliştirmek için YZ ile ilgili araçların potansiyelini göstermesidir (Martínez - Comesaña vd., 2023).

Lu vd. (2024), Analitik Hiyerarşi Süreci (AHP) ile LLM'lerin değerlendirme becerisini birleştirerek öğrencilerin açık uçlu sorulara verdikleri yanıtların çok kriterli analizini yapabilmeyi sağlayan yeni bir yöntem önermektedir. İki aşamadan oluşan bu yöntem, insan yargılarına yakın değerlendirmeler elde etmeyi amaçlamaktadır. LLM'ler aracılığı ile farklı yanıt çiftlerinin karşılaştırıldığı ilk aşamada, en sık tekrar eden değerlendirme ölçütleri belirlenmiş; bu yanıtlardan neden birinin diğerinden daha iyi olduğu açıklanarak ölçütler sıralanmıştır (dil kullanımının doğruluğu, gerekçelendirme, netlik gibi). Bu açıklamalar analiz edilerek en sık tekrarlanan ve önemli görülen nedenler, yani değerlendirme kriterleri oluşturulmuştur. AHP sürecinin olduğu ikinci aşamada ise LLM aracılığı ile belirlenmiş; her bir ölçüte göre yanıt çiftleri tekrar karşılaştırılmış ve nihai puanlama yapılmıştır. Her birinde yaklaşık seksen kişinin yanıtlarının değerlendirildiği, açık uçlu sorulardan oluşan dört farklı veri seti kullanılarak gerçekleştirilen çalışmada bu yöntemin, klasik puanlamaya göre daha esnek ve açıklanabilir bir değerlendirme süreci sunduğu; daha yüksek doğruluk ve tutarlılık sağladığı belirtilmektedir.

Chen vd. (2024), IU International University of Applied Sciences'ta yer alan lisans ve yüksek lisans öğrencilerinin sınav sorularına verdikleri yanıtların bulunduğu geniş bir veri setini kullanmışlardır. Farklı derslerdeki sınav soruları ve öğrenci yanıtlarını kullanarak eğitilmiş bir açık kaynak dönüştürücü model (transformer-based model) ile kısa cevapların otomatik puanlanmasını gerçekleştirmişlerdir. Bu modelin temel amacı, insan puanlayıcılarda görülen tutarsızlıkların ve öznel yargıların yer almadığı, güvenilir bir değerlendirme sistemi oluşturmaktır. Çalışmaya göre model, kişilerin verdiği puanlar ile karşılaştırıldığında hem tutarlı hem de doğru değerlendirmeler yaparak üstün bir performans göstermiştir. Ayrıca, sistem açıklanabilirlik açısından incelenmiş ve verilen puanlar ile ilgili kararlarının hangi nedene dayandığı açıkça analiz edilmiştir. Özellikle

farklı ders alanlarındaki sorular ve farklı öğrenci yanıtlarını içeren bir veri seti kullanılması sistemin genellenabilirliği açısından önemlidir. Bunun yanında çalışmada, etik sorumluluklar, önyargılar ve öğrencilere verilecek geri bildirimler de tartışılmıştır. Elde edilen sonuçlara bakarak, eğitim alanında YZ sistemlerinin insan merkezli değerlendirmeden daha objektif ve ölçeklenebilir bir sistem olduğu vurgulanmıştır.

Zhao vd. (2025), LLM'lerin eğitilmeden ya da özel ayarlamalar olmadan, yalnızca birkaç örnek gösterilerek (few-shot: YZ sistemlerinin çok az sayıda örnek verildikten sonra yeni bir görev yapabilme yeteneği) öğrenci çalışmalarını değerlendirebilme ve puanlama becerisini ortaya koymuştur. LLM'lerin açık uçlu sorulara öğrencilerin verdiği yanıtları değerlendirme yeteneklerini inceleyen çalışmada, GPT-4, GPT-4o ve o1-preview sistemlerinin az sayıda örnekle kısa cevaplı soruları değerlendirme performansları karşılaştırılmıştır. Araştırmada puanlama ölçütlerinin açıkça yazdığı yönergeler kullanılarak modellerin tutarlılığı ve doğruluğunda artış gözlemlenmiş; ayrıca öğretmenler tarafından puanlanan örnekler modele sunulurken, puanlama performanslarının arttığı belirlenmiştir. Sonuç olarak bu çalışma ile LLM'lerin az örnek paradigması altında bile insan değerlendirecilerle yakın performans sergilediği ortaya konmuş ancak modellerin başarısının çeşitli faktörlere bağlı olduğu belirtilerek (örneğin seçilen yanıtların temsil gücü, değerlendirme ölçütlerinin açıklığı ve ayrıntılı olması, yanıtların zorluğu gibi), sistemlerin eğitimcilerin iş yükünü azaltma potansiyeli taşıdığı, öğretmenlerin yerini almak üzere değil, onlar tarafından yönlendirilerek değerlendirme sürecine yardımcı olmak amacıyla kullanılabileceği ortaya konmuştur.

Alan yazındaki çalışmalar incelendiğinde, ulusal araştırmaların daha çok YZ'nin eğitimde potansiyel kullanımını, teorik çerçevesini ve temel olarak hangi alanlarda kullanılabileceğini keşfetmeye yönelik olduğu görülmektedir. Özellikle akıllı öğretim sistemleri, bireyselleştirilmiş öğrenme yaklaşımları, öğrencilerin kişisel özelliklerinin belirlenmesi ve YZ'nin gelecekte eğitimdeki yerine odaklanılmış; ancak somut uygulama geliştirme ya da deneysel çalışmaların sınırlı olduğu gözlemlenmektedir. Ancak uluslararası araştırmaların, özellikle son yıllarda LLM tabanlı değerlendirme sistemlerinin geliştirilmesi, açık uçlu soruların otomatik puanlanması ve çok kriterli değerlendirme yöntemlerinin uygulanması gibi daha özel, belirgin ve teknik konulara yoğunlaştığını göstermektedir. Uluslararası çalışmalar, YZ sistemlerinin eğitimdeki kullanımını pratik uygulamalar ve beceri temelli değerlendirmeler üzerinden ele almış; bu teknolojilerin gerçek eğitim süreçlerindeki etkinliğini güncel verilerle desteklemiştir. Tüm çalışmaların ortak sonucu olarak, hem ulusal hem uluslararası literatürde YZ'nin

öğretmenlerin yerini alacak bir alternatif olmadığı, eğitim uygulamalarını destekleyici bir araç olduğu sonucu çıkmaktadır. İnsan-makine işbirliği modeli çerçevesinde, öğretmenlerin pedagojik uzmanlığının YZ sistemlerinin desteği ile güçlendirilmiş hibrit bir yaklaşımla eğitime entegre edilmesi gerektiği sonucuna varılmaktadır.



BÖLÜM 3

YÖNTEM

Bu bölümde araştırmanın modeli, evren ve örnekleme, veri toplama süreci ve toplanan verilerin çözümlenmesi ile ilgili bilgilere yer verilmiştir.

Araştırmanın Modeli

Bu araştırmada; 7. sınıf matematik dersinde sınıf içi ölçmede kullanılan yazılı yoklama sınavlarındaki açık uçlu sorulara verilen yanıtları dereceli puanlama anahtarı ile puanlayan öğretmenin verdiği puanlar ile aynı anahtara bağlı olarak puanlama yapmak için eğitilmiş YZ puanlamaları; sınavın test ve madde istatistikleri ve öğrencilere yapılan bireyselleştirilmiş geri bildirimler karşılaştırılmıştır. Çalışmada YZ tabanlı değerlendirme sistemleri ile öğretmen değerlendirmeleri arasındaki tutarlılık, objektiflik ve entegrasyon süreçlerini incelemek amacıyla açıklayıcı sıralı karma desen (Creswell, 2014) kullanılmıştır. Araştırma, nitel ve nicel veri toplama tekniklerini birleştirerek kapsamlı ve derinlemesine bir analiz sunmayı hedeflemektedir. İlk aşamada nicel veriler toplanıp analiz edilmiş, ardından nitel aşama ile bulgular derinleştirilmiştir. Bu yaklaşım, YZ ve öğretmen değerlendirmeleri arasındaki ilişkileri ve ayrışmaları tespit etmek ve bunların nedenlerini keşfetmek için tercih edilmiştir.

Çalışma Grubu

Bu çalışmada, bir özel okulun üç farklı ildeki beş farklı yerleşkesinden (Ankara Oran, Ankara İncek, Antalya Muratpaşa, Antalya Manavgat, Sakarya) 2023 - 2024 akademik yılının ikinci döneminde öğrenim gören ortaokul 7. sınıf öğrencileri ve aynı okullarda görev yapan öğretmenlerden kıdemli matematik öğretmenleri yer almaktadır. Öğretmenlerin puanlama konusundaki yetkinliklerini artırmak amacıyla MEB tarafından verilen hizmet içi eğitim programlarına katılımı sağlanmıştır. Puanlayıcı YZ olarak *Claude (Sonnet)* ile oluşturulmuş *chatbot* az sayıda örnek (few-shot learning) ve dereceli puanlama anahtarı gösterilerek eğitilmiştir.

Çalışma grubu, amaçlı örnekleme yöntemlerinden tipik durum örnekleme ile seçilmiş olan 217 7. sınıf öğrencisi, dört ölçme uzmanı ve yazılıları puanlamış olan beş matematik öğretmenin de dahil olduğu otuz altı öğretmenden oluşmaktadır. Öğrencilerin 110'u (%50.7) kız, 107'si (%49.3) erkek öğrencilerden oluşmaktadır. Öğrencilerin yaş dağılımı incelendiğinde, 189'unun (%87.1) 2011 doğumlu, 20'sinin (%9.2) 2010 doğumlu ve 8'inin (%3.7) 2012 doğumlu olduğu görülmektedir. Araştırmaya katılan öğrencilerin büyük çoğunluğu (%87.1) yazılı sınavı uygulamasına katıldıkları 2024 yılında 13 yaşında olup, tipik bir 7. sınıf öğrenci profilini yansıtmaktadır. Çalışma grubu cinsiyet dağılımı açısından dengeli bir kompozisyona sahiptir. Araştırmanın gerçekleştirildiği özel okulun yerleşkelerine göre öğrenci dağılımı şu şekildedir: Antalya yerleşkesinden 42 (%19.4), Diyarbakır yerleşkesinden 24 (%11.1), Sakarya yerleşkesinden 10 (%4.6), Ankara Oran yerleşkesinden 88 (%40.5) ve Ankara İncek yerleşkesinden 53 (%24.4) öğrenci çalışmaya katılmıştır. En yüksek katılım Ankara'daki iki yerleşkeden (toplam %64.9) sağlanırken, en düşük katılım Sakarya yerleşkesinden (%4.6) gerçekleşmiştir.

Geri bildirimlerin değerlendirilmesi için sınava katılan öğrenciler yüksek puandan düşük puana sıralanarak 100-80, 79-50 ve 50-0 puan aralıkları ile üst, orta ve alt seviyeler belirlenmiştir. Her bir seviyeden on öğrenci rastgele seçilerek odak gruplar oluşturulmuştur. Açık uçlu soruların ve dereceli puanlama anahtarının geliştirilmesi aşamalarında beş matematik öğretmeni ve ölçme ve değerlendirme alan uzmanlarından dört kişi yer almıştır.

YZ sistemlerinin eğitim sistemine mevcut durumdaki entegrasyonu ve gelecekteki kullanımı ile ilgili olarak çevrimiçi ankete 36 öğretmen katılmıştır. Anketi yanıtlayan öğretmenlerin branş dağılımına bakılmış ve anket verilerinin karşılaştırılması amacı ile sayısal, sözel, yabancı dil ve diğer branşlar şeklinde gruplar oluşturulmuştur. Yazılıları puanlamış beş öğretmenin de içinde olduğu matematik branşından toplam altı öğretmen (%16.7) ve fen bilimleri dersinden katılan üç öğretmen (%8.3) ile sayısal grup; altı Türkçe öğretmeni (%16.7) ve iki sosyal bilgiler öğretmeni ile (%5.6) sözel grup; farklı yabancı dil alanlarından yedi öğretmen (%19.4) ile yabancı dil grubu; dört sınıf öğretmeni (%11.1), beş psikolojik danışman (%13.9), iki ölçme ve değerlendirme uzmanı (%5.6) ve bir bilişim teknolojileri öğretmeni (%2.8) ile diğer branşlar grubu belirlenmiştir. Nihai olarak araştırma sürecine katılan toplam 217 öğrenci (başarı seviyesine göre seçilen 30 öğrenci dahil), 36 öğretmen (yazılıları puanlayan beş matematik öğretmeni dahil) ve 4 ölçme ve değerlendirme uzmanı bulunmaktadır. Katılımcıların sayıları ve kullanılan uygulama detayları ile ilgili bilgiler Tablo 1'de sunulmuştur.

Tablo 1

Katılımcı Bilgileri

Katılımcılar	Sayı	Uygulama Detayı
7. sınıf öğrencileri	217 öğrenci	Tipik durum örnekleme ile seçilmiş, 2023-2024 akademik yılı ikinci dönemi öğrencileri Üst-orta-alt seviyelerden rastgele örnekleme ile seçilmiş, 2023-2024 akademik yılı ikinci döneminden 30 öğrenci
Claude (Sonnet)	YZ modeli	Dereceli puanlama anahtarı ve few-shot learning ile eğitilmiş
Öğretmenler	36 öğretmen	Çevrimiçi anketi yanıtlayan öğretmenler
Ölçme ve değerlendirme uzmanları	4 uzman	Açık uçlu soru ve puanlama anahtarı geliştirme sürecinde

Verilerin Toplanması

Bu çalışmada, veri toplama aracı olarak nicel veriler için yazılı sınav ve nitel veriler için görüşme soruları kullanılmıştır. Araçlar hakkında bilgiler aşağıda sunulmuştur.

Veri Toplama Araçları

Araştırmada, araştırma soruları doğrultusunda dört ana kategoride veri toplanmıştır. Birinci veri toplama aracı, 7. sınıf matematik dersine yönelik açık uçlu soruları içeren yazılı yoklama sınavıdır (EK. 1). Bu sınav, ölçme ve değerlendirme alanında uzman dört kişi ile matematik öğretmenlerinin işbirliğiyle 7. sınıf matematik dersi öğretim programındaki konu kazanımlarını kapsayan bir belirtke tablosu doğrultusunda hazırlanmıştır.

Bu bağlamda araştırmanın yapıldığı okul bünyesinde, yazılı sorularının kapsamı için MEB tarafından yayınlanmış olan 7. sınıf matematik dersi ikinci dönem birinci ortak yazılıları senaryolarından 6. senaryo seçilmiştir. Senaryoya bağlı olarak konu kapsamına uygun şekilde kazanım ilişkilendirmesi yapılmış; senaryoda belirtilen sayı kadar açık uçlu maddeden oluşan matematik yazılısı matematik öğretmenleri tarafından hazırlanmıştır. Yazılı sınavı soruları, kapsam geçerliği açısından güçlü bir yapı sergilemektedir. Milli Eğitim Bakanlığı öğretim programında yer alan "Cebir", "Sayılar ve İşlemler" öğrenme

alanlarına ait toplamda on kazanımı içermekte olup, her bir kazanıma yönelik birer madde içerecek şekilde dengeli bir dağılım göstermektedir. Belirtke tablosundaki kazanımların Bloom taksonomisine göre incelenmesi, testin hatırlama basamağından başlayarak anlama, uygulama, analiz ve değerlendirme düzeylerine kadar çeşitli bilişsel süreçleri kapsadığını ortaya koymaktadır. Bu hiyerarşik yapı, testin farklı düşünme becerilerini kapsamlı bir şekilde ölçtüğünü göstermektedir. Bu bağlamda, hazırlanan ölçme aracının hedef kitlenin matematik başarısını geçerli ve güvenilir bir şekilde ölçebilecek nitelikte olduğu söylenebilir. Soru sayısı ve konu içeriği bakımından öğretmenler ve uzmanlar tarafından uygun bulunan senaryonun kazanım dağılımı Tablo 2’de sunulmuştur.

Tablo 2

7. Sınıf Matematik Dersi 2. Dönem 1. Yazılı Konu Soru Dağılım Tablosu

Öğrenme Alanı	Alt Öğrenme Alanı	Kazanımlar	6. Senaryo (10 soru)
CEBİR	Eşitlik ve Denklem	M.7.2.2.2. Birinci dereceden bir bilinmeyenli denklemi tanır ve verilen gerçek hayat durumlarına uygun birinci dereceden bir bilinmeyenli denklem kurar.	1
		M.7.2.2.3. Birinci dereceden bir bilinmeyenli denklemleri çözer.	1
		M.7.2.2.4. Birinci dereceden bir bilinme yenli denklem kurmayı gerektiren problemleri çözer	1
		M.7.1.4.2. Birbirine oranı verilen iki çokluktan biri verildiğinde diğerini bulur.	1
SAYILAR VE İŞLEMLER	Oran ve Orantı	M.7.1.4.4. Doğru orantılı iki çokluk arasındaki ilişkiyi ifade eder.	1
		M.7.1.4.6. Gerçek hayat durumlarını inceleyerek iki çokluğun ters orantılı olup olmadığına karar verir.	1
		M.7.1.4.7. Doğru ve ters orantıyla ilgili problemleri çözer.	1
	Yüzdeler	M.7.1.5.1. Bir çokluğun belirtilen bir yüzdesine karşılık gelen miktarını ve belirli bir yüzdesi verilen çokluğun tamamını bulur.	1
		M.7.1.5.3. Bir çokluğu belirli bir yüzde ile arttırmaya veya azaltmaya yönelik hesaplamalar yapar.	1
		M.7.1.5.4. Yüzde ile ilgili problemleri çözer.	1

Yazılı sınavında yer alan açık uçlu maddeler için kapsamlı bir değerlendirme anahtarı geliştirilmiştir. Bu anahtarda her madde için tam doğru yanıtlar ve kabul edilebilecek kısmen doğru yanıtlar detaylı olarak tanımlanmış, her bir yanıt düzeyi için

verilecek puanlar belirlenmiştir. Özellikle problem çözme ve analiz gerektiren üst düzey kazanımlarda, öğrencilerin farklı çözüm stratejilerini ve matematiksel muhakeme süreçlerini değerlendirebilmek amacıyla esnek bir puanlama yaklaşımı benimsenmiştir. (Tablo 3).

Tablo 3

Örnek Açık Uçlu Madde Dereceli Puanlama Anahtarı

Tam Doğru Yanıt
10
Kısmen Doğru Yanıtlar
20
21
22
Yanlış Yanıtlar
30
Boş Yanıt
40
Diğer Yanıtlar
50

Yazılı yoklama sınavlarının değerlendirilmesi sonucunda elde edilen test ve madde istatistikleri, YZ ve öğretmen değerlendirmelerinin güvenilirlik, ayırt edicilik ve zorluk düzeyi gibi psikometrik özelliklerini karşılaştırmak amacıyla kullanılmıştır. İkinci veri toplama aracı, öğrencilere YZ ve öğretmenler tarafından sağlanan bireysel geri bildirimlerin öğrenciler açısından yorumunu belirlemek amacıyla yapılan odak grup görüşmeleri için kullanılan yarı yapılandırılmış görüşme formudur. Bu yorumlar alt, orta ve üst grup öğrencilerinden seçilen odak gruplarla yüz yüze görüşme şeklinde yapılmış, görüşme esnasında önceden belirlenmiş sorular sorularak öğrenci yanıtları not edilmiştir. Öğrencilere yöneltilen üç ana soru şu şekildedir: i) YZ tabanlı bir sistemden aldığınız geri bildirimler, öğretmenlerinizden aldığınız geri bildirimlere kıyasla ne kadar faydalı? Hangisi daha yararlı oluyor?, ii) YZ tabanlı değerlendirme sistemleri size göre adil mi? Sizi doğru bir şekilde değerlendiriyor mu? Bu konuda herhangi bir endişeniz var mı?, iii) YZ tabanlı bir değerlendirme sistemi sizce ne kadar motive edici? Bu sistem, öğretmenlerin yaptığı değerlendirmelere göre daha mı az, daha mı çok motive edici? şeklinde yarı yapılandırılmış sorular kullanılmıştır. Öğrencilerin verdikleri yanıtlar içerik analizi ile tablollaştırılarak bulgular bölümünde sunulmuştur.

Üçüncü veri toplama aracı ise, üç farklı ildeki özel okul yerleşkelerinde görev yapan beş matematik öğretmeniyle gerçekleştirilen yarı yapılandırılmış görüşmeler ve öğretmen algılarını belirlemeye yönelik kullanılan ankettir. Görüşmeler ve anket, YZ destekli değerlendirme sistemlerinin entegrasyonu, algılanma biçimleri ve süreçte karşılaşılan teknik ve pedagojik zorlukları ortaya çıkarmak amacıyla kullanılmıştır. Yazılı sınavın uygulandığı özel okulun ölçme uzmanları ve matematik öğretmenleri ile bire bir yapılan görüşmelerde kullanılan sorular aşağıda sunulmuştur: i) YZ tabanlı değerlendirme sistemlerini daha önce kullandınız mı? Eğer kullandıysanız, nasıl bir deneyim yaşadınız?, ii) Mevcut eğitim sisteminde YZ tabanlı değerlendirmelerin yeri ve önemi hakkında ne düşünüyorsunuz? Bu sistemlerin öğrenci değerlendirmesindeki rolünü nasıl görüyorsunuz?, iii) Sizce YZ tabanlı değerlendirme sistemleri öğrenci başarısını ölçmede yeterli ve doğru sonuçlar verebiliyor mu? Bu konuda herhangi bir kaygınız var mı?, iv) YZ tabanlı değerlendirme sistemleri, öğretmen değerlendirmeleri ile nasıl karşılaştırılabilir? Sizce hangi yöntemin avantajları ve dezavantajları vardır?, v) YZ tabanlı değerlendirme sistemlerinin derslere entegre edilmesi konusunda ne düşünüyorsunuz? Bu sistemlerin sizin iş yükünüzü nasıl etkilediğini düşünüyorsunuz?, vi) YZ tabanlı değerlendirme sistemlerinin öğrencilere geri bildirim sağlamadaki yeteneği hakkında ne düşünüyorsunuz? Bu geri bildirimlerin öğrencilerin başarısı ve motivasyonu üzerindeki etkileri nelerdir?, vii) Pedagojik açıdan YZ tabanlı değerlendirme sistemlerini kullanmak zorlayıcı oluyor mu? Bu sistemlerin pedagojik uyumluluğu hakkında ne düşünüyorsunuz?, viii) Bu sistemlerin geliştirilmesi için neler önerirsiniz? YZ tabanlı sistemlerin daha etkili hale gelmesi için hangi özelliklerin eklenmesi gerektiğini düşünüyorsunuz?, ix) YZ tabanlı değerlendirme sistemleri ile daha çok çalıştığınızı veya daha az çalıştığınızı düşünüyor musunuz? Bu sistemler, sizin için daha motive edici mi yoksa daha zorlayıcı mı?

Öğretmenlerin, YZ'nin eğitime entegrasyonu, günümüz eğitimindeki yeri ve gelecekte nasıl olması gerektiği ile ilgili görüşlerini öğrenmek için 4'lü Likert tipinde oluşturulan 10 soruluk anketin soruları şu şekildedir: i) YZ tabanlı puanlama sistemlerinin mevcut eğitim süreçlerine entegrasyonu başarılıdır. ii) YZ destekli puanlama sistemlerinin kullanılması öğretim ve değerlendirme süreçlerini iyileştirir. iii) YZ ile yapılan değerlendirmeler öğretmen değerlendirmesi kadar iyi değildir. iv) YZ destekli puanlama sistemleri eğitimcilerin iş yükünü azaltır. v) YZ destekli puanlama sistemleri her okulda kullanılamayacağı için eğitim alanına entegre edilmemelidir. vi) YZ destekli sistemler öğrencilere etkili geri bildirimler sağlar. vii) YZ tabanlı değerlendirme

sistemlerinin kullanımı zordur. viii) YZ destekli puanlama sistemleri eğitimde daha etkin kullanılabilmesi için teknik ve pedagojik iyileştirmeler gereklidir. ix) YZ destekli puanlama sistemleri eğitimde kullanılmalıdır. x) Bu teknolojilerin eğitim alanında geniş çapta kabul görmesi gereklidir. Anket maddeleri, YZ destekli değerlendirme sistemlerine yönelik literatürde belirlenen temel tartışma konuları (entegrasyon zorlukları, öğretmen rolü, iş yükü azaltma, pedagojik iyileştirmeler) dikkate alınarak araştırmacı tarafından geliştirilmiştir. Kapsam geçerliği için uzman görüşüne başvurulmuştur.

Ek olarak, öğrencilerle yapılan odak grup görüşmelerinde kullanılan sorular şunlardır: i) YZ tabanlı bir sistemle daha fazla çalışmak mı istersiniz yoksa öğretmenlerinizin değerlendirmelerini tercih eder misiniz? Neden?, ii) YZ tabanlı değerlendirme sistemi kullanıldığında kendinizi daha özgür mü yoksa daha kısıtlanmış mı hissediyorsunuz?, iii) Bu sistemleri kullanarak kendi ilerlemenizi daha kolay takip edebileceğinizi düşünüyor musunuz?, iv) YZ tabanlı değerlendirmeleri daha fazla kullanmak ister misiniz? Neden?

Veri Toplama Yöntemi

Yazılı sınavları ve puanlama anahtarları ilgili öğretmenlere iletilerek sınıflarında sınavların uygulanması sağlanmıştır. Tüm öğrencilerin yanıtladığı yazılı kağıtları tarayıcı ile taranıp dijital ortama aktarıldıktan sonra her okuldaki görevli öğretmen tarafından o yerleşkede yazılıya girmiş olan öğrencilerin yazılıları puanlama anahtarına uygun şekilde puanlanmış ve sonuç tabloları bir araya getirilmiştir. YZ sisteminde az örnekle ve dereceli puanlama anahtarı ile eğitilen *chatbota* taratılmış öğrenci kağıtları yüklenerek puanlama anahtarına uygun şekilde değerlendirmesi istenmiştir.

Sonuçlar değerlendirildikten sonra YZ tabanlı değerlendirme sistemlerinin ve öğretmenlerin geri bildirimleri öğrencilerle paylaşarak yazılı notlarını öğrenmeleri ve hata yaptıkları kazanımları fark etmeleri sağlanmıştır. Gruplandırma yapılırken öğretmen puanları ve YZ puanlarının karşılaştırılmasında yüzdelik dilim yöntemi (%27-%46-%27) yerine puan aralığı yöntemi (80-100, 50-79, <50) kullanılmasının daha tutarlı sonuçlar verdiği görülmüş; öğretmen puanlamasında yazılı notları büyükten küçüğe sıralanarak 100 ila 80 puan arası üst grup, 79 ila 50 puan arası orta grup ve 50 puan altı alt grup olarak sınıflandırılmıştır. Geri bildirimler ile ilgili görüşme yapılan odak gruplar her seviyeden rastgele 10 öğrenci olarak seçilmiştir. YZ sistemlerinin mevcut eğitim süreçlerine entegrasyonu ve bu süreçte karşılaşılan zorlukların analiz edilebilmesi için YZ destekli

sistemler ile ilgili sorulardan oluşan bir anket hazırlanmış; yazılıların yapıldığı okullarda süreci yaşamış olan öğretmenlere ve ölçme ve değerlendirme uzmanlarına uygulanmıştır.

Verilerin Çözümlemesi

Bu çalışmada, açık uçlu sınav sorularının değerlendirilmesinde insan puanlayıcılar ile YZ destekli sistemlerden *Claude (Sonnet)*'un performansları karşılaştırılmıştır. Öğretmen ve YZ puanları arasındaki tutarlılık ve farklılıklar, betimsel istatistikler kullanılarak analiz edilmiştir. Testlerin sonuçları, insan puanlayıcılar ve YZ destekli robotların değerlendirmeleri tablolar halinde eklenerek sonuçları karşılaştırılmıştır. Tabloların oluşturulması ve istatistiklerin hesaplanması için Microsoft Excel (Microsoft Corporation, 2021), SPSS 26 (IBM Corporation, 2019) ve R programı (R Core Team, 2023) psych paketi (Revelle, 2023) kullanılmıştır. Analizlerde, puanlamalar arasındaki uyum, puan dağılımları ve varyanslar incelenmiş; iki puanlayıcı arasındaki farklılıkların anlamlılığı değerlendirilmiştir.

Çalışmanın 1. alt amacı olan YZ ve öğretmen puanlamaları arasında tutarlılık ve objektiflik düzeylerinin karşılaştırılması amacıyla; testin iç tutarlılığı için Cronbach Alfa katsayısı hesaplanmıştır. Öğretmen ve YZ puanlarında madde güçlük ve ayırt edicilik indekslerinin uyumu için, normal dağılım varsayımının sağlanmaması nedeni ile Spearman Brown sıra farkları korelasyonu katsayısı ve farkların anlamlılığı için Wilcoxon İşaretli Sıralar Testi kullanılmıştır. Ayrıca ağırlıklandırılmış kappa katsayısı (QWK) kullanılarak insan puanlayıcı ve YZ destekli sohbet robotu değerlendirmelerinde ortalama puanlar ve sorulara verilen tekil puanlardaki farkların anlamlılığına bakılmıştır. Puanlayıcılar arasındaki uyum analizinin güvenilirliği için “tam puan uyumu” ve “büyük puan farklılıkları” incelenmiştir. Ağırlıklandırılmış Kappa katsayısının formülü şu şekildedir:

$$QWK = 1 - \frac{\sum_i \sum_j w_{ij} \times p_{ij}}{\sum_i \sum_j w_{ij} \times e_{ij}}$$

w_{ij} = i ve j kategorileri arasındaki ağırlık (anlaşmazlık derecesini gösteren)

p_{ij} = i ve j kategorilerinin gözlemlenen oranı

e_{ij} = i ve j kategorilerinin beklenen oranı (şans eseri uyum)

Kappa katsayısının değer aralıkları ve bu aralıkların anlamlarına ilişkin belirlenen ölçütler Tablo 4'te yer almaktadır (Landis ve Koch, 1977).

Tablo 4

Kappa İstatistiğinin Yorumlanmasına İlişkin Değer Aralıkları

QWK Değeri	Anlamı
0,00-0,20	Zayıf
0,21-0,40	Hafif
0,41-0,60	Orta
0,61-0,80	Güçlü
0,81-1	Mükemmel uyum

Tüm puanların karşılaştırılmasının yanında tam puan ve kısmi puan uyumları da değerlendirmelerin güvenilirliği için kullanılmaktadır. Gwet (2014) "Handbook of Interrater Reliability" kitabında, değerlendiriciler arasındaki uyum analizinin genel uyum istatistikleriyle sınırlandırılmaması gerektiğini; daha kapsamlı bir değerlendirme için "tam puan uyumu" gibi değerlendiricilerin birebir aynı puanları verdiği durumları ve "büyük puan farklılıkları" gibi değerlendiriciler arasında önemli derecede görüş ayrılığı olan durumları incelemenin de kritik önemli olduğunu savunmaktadır. Bu amaçla tam puan uyumu ve büyük puan farklılıkları hesaplanmıştır.

Çalışmanın 2. alt amacı olan her öğrenciye özel geri bildirim verilmesi sürecinde, YZ'nin ve öğretmenin sağladığı geri bildirimlerin farkları ve benzerlikleri analiz edilmiştir. Öğretmenin ve YZ'nin aynı öğrencilere verdiği geri bildirimlerin içeriklerine yönelik görüşme soruları, öğrenci başarı düzeylerine göre oluşturulmuş odak grup görüşmesinde öğrencilere yöneltilmiştir. Odak gruplar oluşturulduktan sonra başarı düzeylerine göre oluşturulan gruplarda öğretmen puanlaması ile YZ puanlaması arasındaki ilişkinin gücünü belirlemek amacıyla Spearman Brown sıra farkları korelasyon analizi uygulanmıştır. Analiz sonucunda elde edilen korelasyon katsayıları, Cohen (1988) kriterleri doğrultusunda yorumlanmıştır (0.10-0.29 zayıf, 0.30-0.49 orta, 0.50 ve üzeri güçlü korelasyon). Sonuçlar içerik analizi kullanılarak kod, kategori ve temalar yardımıyla sunulmuştur.

Çalışmanın 3. alt amacı olan YZ tabanlı değerlendirme sistemlerinin eğitim ortamlarında kullanılma yolları ve amaçlarının belirlenmesinde; 4. alt amacı olan öğretmenler ve öğrencilerin YZ tabanlı değerlendirme sistemlerinin eğitim süreçlerine entegrasyonunu değerlendirmelerinde ve 5. alt amacı olan YZ tabanlı değerlendirme sistemlerinin entegrasyonu sürecinde karşılaşılan teknik ve pedagojik zorlukların belirlenmesinde öğretmen görüşmeleri, öğretmen anketi ve öğrenciler ile odak grup görüşmesi sorularının yanıtları kullanılmıştır.

Anket verilerinin betimsel analizlerinde, Likert ölçeğinin sıralı veri yapısına sahip olması nedeniyle, merkezi eğilim ölçüsü olarak medyan (ortanca) değerleri, dağılım ölçüsü olarak ise çeyreklikler arası açıklık (IQR - Interquartile Range) hesaplanmıştır. IQR, 75. percentil (Q3) ile 25. percentil (Q1) arasındaki fark olarak tanımlanmış olup, verilerin merkezi %50'lik diliminin dağılım genişliğini göstermektedir. Düşük IQR değerleri katılımcılar arasında fikir birliğini, yüksek IQR değerleri ise görüş farklılığını işaret etmektedir. Ankete katılan öğretmenlerin branş bazında karşılaştırmaları için sıra ortalaması (rank mean) analizi uygulanmıştır. Sıra ortalaması, parametrik olmayan veriler için gruplar arası karşılaştırma yapmaya olanak sağlayan kararlı bir istatistiksel yöntemdir (Büyüköztürk, 2017). Görüşme sorularında içerik analizi kullanılarak kod, kategori ve temalar yardımıyla elde edilen sonuçlar sunulmuştur. Veri çözümleme için kullanılan yöntemler Tablo 5'te yer almaktadır.

Tablo 5

Veri Çözümleme Yöntemleri

Analiz Yöntemi	Kullanım Amacı	Uygulanan Veriler
Güvenirlilik Analizleri	Yazılı yoklama sınavlarının istatistiksel güvenirliğini belirlemek	Test ve madde istatistikleri
Madde Analizi	YZ ve öğretmen değerlendirmelerine göre soruların ayırt edicilik ve zorluk düzeylerini karşılaştırmak	Yazılı yoklama sınavı maddeleri
Ağırlıklandırılmış Kappa	YZ puanları ile öğretmen puanları arasındaki uyum düzeyini belirlemek	7. sınıf matematik yazılı yoklama puanları
Tematik Analiz	YZ ve öğretmen geri bildirimlerindeki ortak tema ve farklılıkları belirlemek	Odak grup görüşme verileri
Betimsel Analiz	YZ'nin eğitim sürecine eklemlenmesine ilişkin öğretmen algılarını analiz etmek	Anket verileri
Tematik Analiz	YZ'nin değerlendirme performansına ilişkin öğretmen algılarını analiz etmek	Yarı yapılandırılmış görüşme
Tematik Analiz	YZ'nin eğitimdeki yerine ilişkin öğrenci algılarını analiz etmek	Odak grup görüşme verileri

BÖLÜM 4

BULGULAR VE YORUMLAR

Bu bölümde çalışmanın alt amaçlarına yönelik elde edilen veriler tablolar halinde sunulmuş ve bulgular yorumlanmıştır.

Matematik Yazılı Sınavının Güvenirlik ve Madde Özellikleri

Bu bölümde, 7. sınıf matematik yazılı sınavının psikometrik özellikleri incelenmiştir. Analizler hem YZ hem de öğretmen değerlendirmeleri için ayrı ayrı gerçekleştirilmiştir. Testin iç tutarlılığını belirlemek için Cronbach Alfa katsayısı, öğretmen ve YZ puanları arasındaki madde güçlük ve ayırt edicilik indekslerinin tutarlılığını değerlendirmek için ise Spearman Brown sıra farkları korelasyon katsayısı hesaplanmış, elde edilen bulgular sırasıyla sunulmuştur. İnsan puanlayıcı ile YZ destekli sohbet robotunun değerlendirmeleri arasındaki farkların anlamlılığı, ortalama puanlar ve tekil soru puanları üzerinden Ağırlıklandırılmış Kappa (QWK) katsayısı ile analiz edilmiştir. Ayrıca, puanlayıcılar arası uyumun güvenirliliğini incelemek amacıyla “tam puan uyumu” ve “büyük puan farklılıkları” dikkate alınmıştır.

Testin Güvenirliğine Dair Bulgular

İlk araştırma sorusu “YZ ve öğretmen puanlamaları arasında tutarlılık ve objektiflik düzeyleri nasıldır?” için kullanılan yazılı sınavların güvenirliliğini test etmek amacıyla Cronbach Alfa iç tutarlılık katsayıları hesaplanmıştır. Öğretmen puanları ve YZ puanları için hesaplanan katsayılar Tablo 6’da sunulmuştur.

Tablo 6

Testin Güvenirlik Katsayıları

Değerlendirme Türü	Cronbach Alfa	Madde Sayısı	Öğrenci Sayısı
Öğretmen Değerlendirmesi	.82	10	217
YZ Değerlendirmesi	.88	10	217

Her iki puanlamada da güvenilirliğin iyi düzeyde olduğu belirlenmiştir. Alanyazında eğitim alanında .70 değerinin üzerindeki güvenilirlik katsayıları kabul edilebilir, .80 üzerindeki değerler ise iyi düzeyde güvenilirlik olarak değerlendirilmektedir (Büyüköztürk, 2021). Öğretmen puanlamalarının Cronbach Alfa katsayısı .82, YZ puanlamalarının Cronbach Alfa katsayısı ise .88 olarak bulunmuştur. YZ puanlamalarının güvenilirlik katsayısının öğretmen puanlamalarına göre daha yüksek olduğu görülmüştür. Bu durum, YZ değerlendirme sisteminin daha tutarlı puanlama yaptığını düşündürmektedir. Her iki puanlamada da yüksek güvenilirlik değerlerinin olması, ölçme aracının güvenilir sonuçlar ürettiğini ve elde edilen bulguların güvenilirliğini desteklemektedir.

Testin bütününe ilişkin hesaplanan güvenilirlik katsayısının yanı sıra, her bir maddenin testin bütünüyle ne ölçüde tutarlı olduğunu gösteren istatistiklerin incelenmesi de, ölçme aracından elde edilen puanların güvenilirliği hakkında önemli bilgiler sunmaktadır. Öğretmen ve YZ puanlamalarının her madde için toplam istatistikleri Tablo 7’de sunulmuştur.

Tablo 7

Öğretmen ve YZ Puanlamasında Madde Toplam İstatistikleri

	Madde Ortalama		Standart Sapma		Madde-Toplam Korelasyonu		Madde Silindiğinde Cronbach Alfa	
	Öğrt.	YZ	Öğrt.	YZ	Öğrt.	YZ	Öğrt.	YZ
M1	8.76	7.41	1.91	3.47	.48	.63	.82	.87
M2	8.59	8.70	2.56	2.29	.70	.70	.80	.86
M3	7.36	6.99	3.61	3.65	.68	.78	.80	.85
M4	9.17	8.47	1.63	2.45	.58	.67	.81	.86
M5	8.88	8.58	2.34	2.35	.55	.73	.81	.86
M6	8.41	8.29	2.99	2.90	.65	.68	.80	.86
M7	7.79	6.87	3.22	3.78	.61	.76	.81	.86
M8	6.24	5.95	3.93	3.75	.67	.67	.81	.87
M9	9.21	9.21	1.99	1.98	.66	.66	.80	.87
M10	8.23	8.24	2.96	2.99	.73	.72	.79	.86

Madde istatistikleri incelendiğinde, YZ aracının genel olarak daha düşük puanlar verdiği dikkat çekmektedir. Öğretmenlerin madde puanlarında ortalamalar 6.24 ile 9.21 arasında değişirken, YZ madde puanlarına bakıldığında bu aralık 5.95 ile 9.21

arasındadır. Özellikle madde 1 (öğretmen: 8.76, YZ: 7.41), madde 7 (öğretmen: 7.79, YZ: 6.87) ve madde 8'de (öğretmen: 6.24, YZ: 5.95) YZ'nin öğretmenlere göre daha düşük puanlar verdiği göze çarpmaktadır. Bu puanlamaların tutarlılığını görebilmek için standart sapma değerlerini incelemek gerekmektedir. Her iki değerlendirme sisteminde de 3. madde (öğretmen: 3.61, YZ: 3.65), 7. madde (öğretmen: 3.22, YZ: 3.78) ve 8. madde (öğretmen: 3.93, YZ: 3.75) en yüksek değişkenliği göstermektedir. Bunun nedeni maddelerin değerlendirme kriterlerinin yeterince net olmaması veya öznel yoruma açık olması olarak yorumlanabilir. Öte yandan, öğretmen puanlamasında 4. madde (1.63) ve 1. madde (1.91) en düşük standart sapma değerine sahipken YZ puanlamasında 9. madde (1.98) ve 2. madde (2.29) en düşük standart sapma değerlerine sahip maddeler olarak öne çıkmaktadır.

YZ puanlamasının madde-toplam korelasyonları incelendiği zaman öğretmen puanlamasına göre daha yüksek ayırt edicilik gösterdiği dikkat çekmektedir. YZ puanlamasında tüm maddelerin korelasyon değerleri .60'ın üzerindedir. Öğretmen puanlamasında 1. maddenin (.48) ve 5. maddenin (.55) korelasyon değerleri diğerlerinden düşük kalmıştır. Özellikle 7. maddede YZ'nin puanlarında madde korelasyonu (.76) öğretmen korelasyonuna (.61) göre belirgin şekilde yüksektir. Madde ile toplam puan arasındaki korelasyonun yüksek olması, maddede daha tutarlı ve sistematik bir değerlendirme yapıldığını göstermektedir. Zhao ve diğerlerinin (2025) çalışmasında da, matematik problemlerinde GPT-4o'nun %86 doğruluk oranı ile tutarlı performans sergilediği ve özellikle tutarlılık değerlendirmelerinde insan puanlayıcılardan daha iyi sonuçlar verdiği belirtilmektedir. Araştırmada, YZ modellerinin aynı sorular için tutarlı sonuçlar üretme konusunda insan değerlendirmesinden daha güvenilir olduğu tespit edilmiştir; bu çalışmanın bulgularıyla uyumludur. Ayrıca, her bir maddenin silinmesi durumunda güvenilirlik değerlerinin değişimi incelendiği zaman; öğretmen puanlamasında 10. maddenin silinmesi durumunda (.79), YZ puanlamasında ise 3. maddenin silinmesi durumunda (.85) güvenilirliğin en fazla düştüğü görülmektedir.

Madde Güçlük İndeksleri

Her bir maddenin güçlük indeksleri, öğrencilerin maddelere verdikleri doğru yanıt oranları hesaplanarak belirlenmiştir. Öğretmen ve YZ puanlamalarının madde güçlük indeksleri kazanım bazında Tablo 8'de sunulmuştur.

Tablo 8

Madde Güçlük İndeksleri (Kazanım Bazında)

Madde	Kazanım	Öğretmen	YZ	Fark
M1	M.7.2.2.2. Birinci dereceden bir bilinmeyenli denklemi tanır ve verilen gerçek hayat durumlarına uygun birinci dereceden bir bilinmeyenli denklem kurar.	.88	.74	.14
M2	M.7.2.2.3. Birinci dereceden bir bilinmeyenli denklemleri çözer.	.86	.87	-.01
M3	M.7.2.2.4. Birinci dereceden bir bilinme yenli denklem kurmayı gerektiren problemleri çözer	.74	.70	.04
M4	M.7.1.4.2. Birbirine oranı verilen iki çokluktan biri verildiğinde diğerini bulur.	.92	.85	.07
M5	M.7.1.4.4. Doğru orantılı iki çokluk arasındaki ilişkiyi ifade eder.	.89	.86	.03
M6	M.7.1.4.6. Gerçek hayat durumlarını inceleyerek iki çokluğun ters orantılı olup olmadığına karar verir.	.84	.83	.01
M7	M.7.1.4.7. Doğru ve ters orantıyla ilgili problemleri çözer.	.78	.69	.09
M8	M.7.1.5.1. Bir çokluğun belirtilen bir yüzdesine karşılık gelen miktarını ve belirli bir yüzdesi verilen çokluğun tamamını bulur.	.62	.59	.03
M9	M.7.1.5.3. Bir çokluğu belirli bir yüzde ile arttırmaya veya azaltmaya yönelik hesaplamalar yapar.	.92	.92	.00
M10	M.7.1.5.4. Yüzde ile ilgili problemleri çözer.	.82	.81	-.01

Tablodaki verilere bakıldığında en büyük farkların 1. madde (.14), 7. madde (.09) ve 4. maddede (.07) olduğu görülmektedir. YZ sistemi bu soruları puanlarken daha düşük puanlar vermiş, daha zor olarak puanlamıştır. 2, 9 ve 10. maddelerde ise değerlendirmelerin çok yakın olduğu dikkat çekmektedir. Puanlamalar arasında fark olmasına rağmen genel güçlük sıralaması büyük ölçüde benzerlik göstermektedir. İki puanlamada da en zor madde 8, en kolay madde 9 olarak çıkmıştır. Öğretmen ve YZ puanlamalarının madde güçlük indeksleri arasındaki ilişkiyi belirlemek için yapılan Spearman Brown sıra farkları korelasyon analizi sonucunda .842 değeri elde edilmiştir. Bu sonuç, iki puanlayıcının madde güçlüklerini değerlendirmede güçlü bir tutarlılık sergilediğini göstermektedir. Her madde için öğretmen ve YZ puanları arasındaki farka Wilcoxon İşaretli Sıralar Testi ile bakılmış ve madde güçlükleri arasında anlamlı fark ($p = .02$) olduğu belirlenmiştir. YZ değerlendirmeleri genel olarak öğretmen değerlendirmelerine göre daha düşük puanlar (daha zor değerlendirme) verme eğiliminde

gibi görünmesine karşın madde güçlüğü, tek başına o madde ile ilgili karara varılması için yeterli bir bilgi değildir; hangi güçlük derecesinde olursa olsun maddenin ayırt ediciliği önemlidir.

Madde Ayırt Edicilik İndeksleri

Madde ayırt ediciliği, ölçülmek istenen özelliğe sahip olan ve olmayanı ayırabilecek düzeyde iyi maddelerin seçilmesi ve bu amaca hizmet etmeyen maddelerin ise testten çıkarılması için kullanılır. Madde güçlük değerleri ile birlikte değerlendirilerek ölçme aracının amaca uygunluğunu belirlemeyi sağlar. Bir maddenin içinde bulunduğu testle arasındaki korelasyona dayalı bir indeks olduğundan -1 ve +1 arasında değer alır (Alıcı, 2014). Yazılı sınavda puanlanan maddelerin öğretmen ve YZ sistemi puanlamasına göre madde ayırt edicilik indeksleri Tablo 9’da sunulmuştur.

Tablo 9

Madde Ayırt Edicilik İndeksleri ve Yorumları

Madde	Öğretmen Ayırt Edicilik İndeksi	Öğretmen Ayırt Edicilik Düzeyi	YZ Ayırt Edicilik İndeksi	YZ Ayırt Edicilik Düzeyi
M1	.39	Çok İyi	.51	Çok İyi
M2	.61	Çok İyi	.64	Çok İyi
M3	.54	Çok İyi	.70	Çok İyi
M4	.51	Çok İyi	.60	Çok İyi
M5	.44	Çok İyi	.67	Çok İyi
M6	.53	Çok İyi	.59	Çok İyi
M7	.47	Çok İyi	.66	Çok İyi
M8	.52	Çok İyi	.55	Çok İyi
M9	.58	Çok İyi	.60	Çok İyi
M10	.64	Çok İyi	.63	Çok İyi

Pozitif değerlerde madde ayırt edicilik indeks değerleri: ".40 ve üstü, çok iyi madde"; ".30 - .39, iyi ama düzeltme yapılabilir"; ".20 - .29, düzeltilmesi ve geliştirilmesi gerekir"; ".19 ve daha düşük, çok zayıf" şeklindedir (Büyüköztürk, 2021). Bu sınıflandırmalara göre araştırmada kullanılan yazılı sınavın hem öğretmen hem de YZ sistemi puanlaması sonuçlarına göre ayırt ediciliği yüksek bir ölçme aracı olduğu ve amaca hizmet ettiği söylenebilir. Maddeler ayrı ayrı incelendiğinde YZ puanlarının neredeyse tüm maddelerde daha yüksek ayırt edicilik değerlerine sahip olduğu; öğretmen puanlamaları ile karşılaştırıldığında en büyük farkların 5. madde (.22), 7. madde (.19) ve

3. maddede (.16) olduğu görülmektedir. 10, 9 ve 2. maddelerin her iki puanlamada da birbirine yakın ve yüksek ayırt edicilik değerlerine sahip oldukları dikkat çekmektedir. Öğretmen ve YZ puanlamalarının ayırt edicilik indekslerini karşılaştırmak için Spearman Brown sıra farkları korelasyon analizi yapılmış ve sonucunda .68 değeri elde edilmiştir. Bu değer, öğretmen ve YZ puanlarının ayırt edicilik sıralamaları arasında orta düzeyde bir tutarlılık olduğunu göstermektedir. Her madde için öğretmen ve YZ verileri arasındaki farka Wilcoxon İşaretli Sıralar Testi ile bakılmış ve madde ayırt edicilikleri arasında anlamlı fark ($p = .01$) olduğu belirlenmiştir. Madde güçlük indeksindeki yüksek tutarlılığa (.84) kıyasla, ayırt edicilik değerlendirmelerinde puanlayıcılar arasında daha fazla farklılık olduğu görülmektedir. Güçlük ve ayırt edicilik indeksleri bir arada değerlendirildiği zaman YZ sisteminin puanlamasında, daha düşük güçlük indeksleri (daha zor değerlendirme) ile daha yüksek ayırt edicilik indeksleri olması YZ'nin maddeleri daha ayırt edici şekilde değerlendirdiği şeklinde yorumlanabilir. Alan yazında bir çalışmada da, açık uçlu soru karmaşıklığının otomatik puanlanmasında büyük dil modellerinin etkinliği araştırılmış ve YZ sistemlerinin değerlendirme davranışlarına ilişkin önemli bulgular ortaya konmuştur (Raz vd., 2025). Araştırmacılar, 10.000'den fazla insan puanlamalı yanıt kullanarak geliştirdikleri LLM tabanlı sistemin insan puanlamaları ile .73 korelasyon gösterdiğini ve maliyet etkinliği, güvenilirlik ve erişilebilirlik açısından önemli avantajlar sağladığını belirtmişlerdir. Çalışmada dikkat çekici bir bulgu olarak, modelin "düşük karmaşıklıkta soruları hafife alma ve yüksek karmaşıklıkta soruları olduğundan fazla değerlendirme" eğilimi gösterdiği tespit edilmiştir. Bu durum, yaratıcılık değerlendirmesinde otomatik puanlama görevleri için model ince ayarı yapılırken karşılaşılan yaygın bir engel olarak değerlendirilmekte ve YZ sistemlerinin daha ayırt edici değerlendirme yapma eğiliminde olduğunu göstermektedir. Bu bulgular, mevcut çalışmadaki YZ puanlamalarının daha düşük güçlük indeksleri ile daha yüksek ayırt edicilik indeksleri göstermesi bulgularıyla paralellik göstermektedir.

YZ ve Öğretmen Değerlendirmelerinin Karşılaştırmalı Analizi

Bu bölümde, açık uçlu soruların değerlendirilmesinde YZ sistemleri ile öğretmenlerin puanlama yaklaşımları arasındaki tutarlılığın incelenebilmesi için ağırlıklandırılmış kappa (QWK) analizi kullanılmıştır. Analiz sonuçları, değerlendirme süreçlerinin güvenilirliğini artırmak ve öğrenci başarısının daha kapsamlı

değerlendirilmesini sağlamak için hangi alanlarda iyileştirmelere ihtiyaç duyulduğuna ışık tutmaktadır.

Ağırlıklandırılmış Kappa Uyum Analizi Sonuçları

Sınava katılmış olan 217 öğrencinin öğretmen ve YZ tarafından değerlendirilmiş olan madde puanları karşılaştırılarak her madde için QWK değeri hesaplanmıştır. Bu hesaplamaya ilişkin madde katsayıları ve öğretmen ile YZ arasındaki anlaşma düzeyleri verileri Tablo 10'da sunulmuştur.

Tablo 10

Değerlendiriciler Arası Madde Puanı Uyum

Madde	Weighted Kappa	Standart Hata	Alt Güven Aralığı	Üst Güven Aralığı	Anlaşma Düzeyi
M1	.35	.10	.17	.53	Hafif
M2	.59	.09	.41	.76	Orta
M3	.65	.06	.53	.77	Güçlü
M4	.39	.12	.17	.62	Hafif
M5	.57	.10	.38	.75	Orta
M6	.52	.09	.35	.69	Orta
M7	.48	.07	.33	.62	Orta
M8	.52	.07	.39	.64	Orta
M9	.64	.11	.44	.85	Güçlü
M10	.62	.08	.47	.77	Güçlü

Tabloda görüleceği üzere, tüm maddeler için kappa değerleri .35 ile .65 arasında değişmektedir, bu da orta ile iyi düzeyde bir uyum olduğunu göstermektedir. Tüm maddelerde standart hataların düşük (.06 - .12) olması tahminlerin güvenilir olduğunu göstermektedir. En yüksek uyum, 3. maddede (.65) görülmektedir. En düşük uyum ise, 1. maddede (.35) gözlenmiştir.

Tüm puanların karşılaştırılmasının yanında "tam puan uyumu" gibi değerlendiricilerin birebir aynı puanları verdiği durumları ve "büyük puan farklılıkları" gibi değerlendiriciler arasında önemli derecede görüş ayrılığı olan durumlar karşılaştırılarak sonuçları Tablo 11'de sunulmuştur. Yüzdelerin yorumları kappa aralıklarına göre uyarlanmıştır.

Tablo 11

Değerlendiriciler Arası Tam Puan Uyumu

Madde	Tam Puan Uyumu (%)	Yorumu
M1	49.8	Orta
M2	64.1	Güçlü
M3	62.2	Güçlü
M4	52.5	Orta
M5	58.5	Orta
M6	66.8	Güçlü
M7	50.7	Orta
M8	47.0	Orta
M9	80.6	Güçlü
M10	70.0	Güçlü

Puanlamalarında büyük farkların (≥ 3) olması durumu da tam puan uyumu kadar önemli bir veridir. Puanlar arası 3 ya da daha fazla fark olan durumların verileri Tablo 12’de sunulmuştur. Yüzdelerin yorumları Gwet’e (2014) göre belirlenmiştir.

Tablo 12

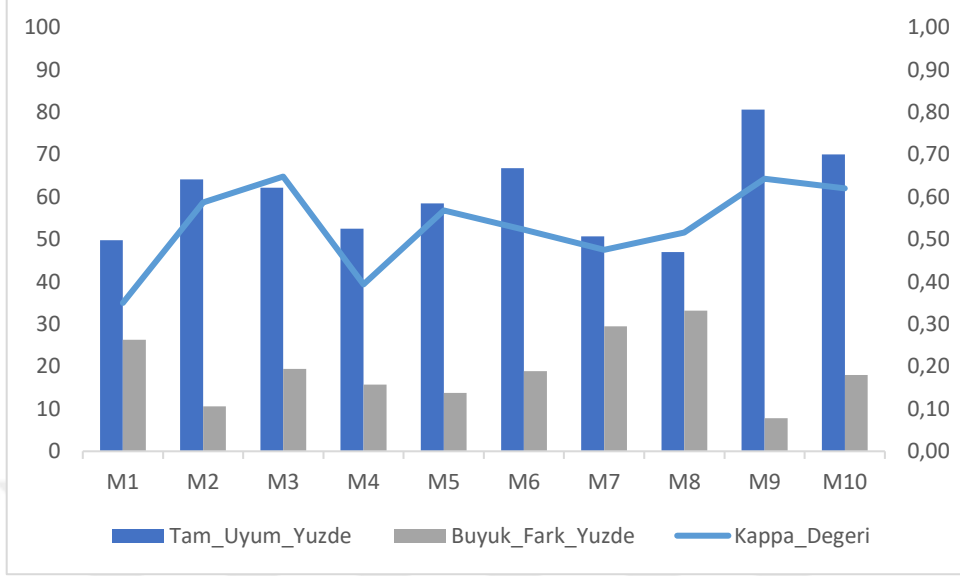
Değerlendiriciler Arası Büyük Puan Farklılıkları

Madde	Büyük Farklılık (%)	Yorumu
M1	26.3	Yüksek düzeyde ayrışma
M2	10.6	Yüksek düzeyde ayrışma
M3	19.4	Yüksek düzeyde ayrışma
M4	15.7	Yüksek düzeyde ayrışma
M5	13.8	Yüksek düzeyde ayrışma
M6	18.9	Yüksek düzeyde ayrışma
M7	29.5	Yüksek düzeyde ayrışma
M8	33.2	Yüksek düzeyde ayrışma
M9	7.8	Orta düzeyde ayrışma
M10	18.0	Yüksek düzeyde ayrışma

Puanlayıcılar arası uyum için yapılan analizlerin karşılaştırmalı gösterimi Grafik 1’de sunulmuştur.

Grafik 1

Madde Bazında Puanlayıcı Uyumu Karşılaştırmaları



Tablolar ve grafik üzerinden maddeler ile ilgili olarak görülen; 9. maddenin (Aşağıdaki soruları yanıtlayınız. a) 80 sayısının % 20'si kaçtır? / b) % 5'i 12 olan sayının tamamı kaçtır?) %80.6'lık tam puan uyumu ve .64 kappa değeriyle en güvenilir madde olduğudur. Bu madde aynı zamanda sadece %7.8'lik büyük fark oranıyla değerlendiriciler arasında yüksek tutarlılık göstermektedir. En büyük ayrışmanın yaşandığı 8. soru (Emre bir kitabın önce %10'unu, daha sonra kalanın %20'sini okuyor. Geriye 288 sayfa kaldığına göre kitap kaç sayfadır?) %47 tam puan uyumu ve %33.2 büyük fark oranıyla en sorunlu madde konumundadır. .52 kappa değeriyle Landis ve Koch (1977) sınıflandırmasına göre "orta" düzeyde bir uyum göstermekte, ancak diğer maddelere kıyasla zayıf kalmaktadır. Alan yazında yapılan bir çalışmada da, soru zorlukları arttıkça YZ modellerinin performans kaybı yaşadığı görülmüştür. Bu çalışmadaki en zor soru olan 8. soru aynı zamanda değerlendiriciler arası tutarlılığın da en düşük olduğu maddedir. Aynı araştırmada sorunun çözümünün karmaşıklığı arttıkça da tutarlılığın azaldığı belirtilmiştir. YZ'nin çok adımlı algortimalarda yetersiz olduğu, mantıksal tutarlılık sorunları ve formal teoremleri ispatlamada eksiklik olduğu ifade edilmiştir (Wang ve diğerleri, 2025).

Genel olarak, tam puan uyumu ile kappa değeri arasında pozitif, büyük fark ile kappa değeri arasında negatif bir ilişki gözlenmektedir. Bu da beklenen bir durumdur ve ölçme aracının iç tutarlılığını desteklemektedir. Değerlendirme sisteminin iyileştirilmesi

için, özellikle 1, 7 ve 8. maddelerin gözden geçirilmesi ya da puanlama kriterlerinin daha net hale getirilmesi gerekmektedir.

YZ ve Öğretmen Geri Bildirimlerinin Nitel Analizi

Bu bölümde, 7. sınıf öğrencilerinin matematik yazılı sınavları sonuçları için öğretmenlerinden ve YZ'den aldıkları geri bildirimler hakkındaki görüşleri nitel veri analizi yöntemiyle incelenmiştir. Yazılı sınava katılmış olan 217 öğrenci not sıralamasına göre üst, orta ve alt gruplar olarak bölünmüş; her gruptan 10'ar öğrenci rastgele seçilerek oluşturulan odak gruplar ile rehberlik dersinde rehber öğretmen eşliğinde görüşme yapılmıştır. Not sıralamasına göre yapılan puan aralığı gruplandırmasında öğretmen ve YZ puanları uyumuna Spearman Brown sıra farkları korelasyonu bakılmış; tüm gruplarda pozitif ve orta düzeyde korelasyonlar elde edildiği (üst grup $\rho = 0.41$, orta grup $\rho = 0.50$, alt grup $\rho = 0.55$) görülmüştür.

YZ'nin ve öğretmenin sağladığı geri bildirimlerin farkları ve benzerlikleri konusunda araştırmanın ikinci alt amacı olan “Öğrencilere verilen bireyselleştirilmiş geri bildirimlerde, YZ'nin ve öğretmenin sunduğu geri bildirimler nasıl farklılık ve benzerlik göstermektedir?” sorusu için oluşturulan odak gruplarda öğrencilere sorular yöneltilmiş ve verilen yanıtlar içerik analizi kullanılarak kod, kategori ve temalar yardımı ile analiz edilmiştir. Bulgular Tablo 13'te sunulmuştur.

Tablo 13

Odak Grup Görüşmesi Geri Bildirim Analizi

GRUP	TEMA	KOD	ÖRNEK CÜMLE	FREKANS
ÜST	Geri Bildirim Kalitesi	Hızlı Erişim	“YZ öğretmenden daha hızlı ve ayrıntılı geri bildirim verebilir”	8/10
		Ayrıntılı Açıklama	“YZ çok ayrıntılı açıklama yapmış ve hatalarımı adım adım göstermiş”	7/10
		Sürekli Ulaşılabilirlik	“YZ'nin geri bildirimleri hemen ve her zaman ulaşılabilir”	6/10
		Tutarlılık	“YZ tarafsız ve tutarlı geri bildirim veriyor”	5/10

(Devam ediyor)

Tablo 13 (Devam)

Odak Grup Görüşmesi Geri Bildirim Analizi

GRUP	TEMA	KOD	ÖRNEK CÜMLE	FREKANS
ÜST	Adalet Algısı	Teknik Karmaşıklık	“Bazen çok teknik oluyor”	4/10
		Önyargısızlık	“Ben insanlardan daha adil olduğumu düşünüyorum çünkü önyargıları yok”	9/10
		Eşit Davranış	“YZ kimseyi tanımıyor herkesi aynı kriterlere göre değerlendiriyor”	8/10
		Objektif Değerlendirme	“YZ tamamen objektif olduğu için çok adil”	7/10
	Motivasyon	Yaratıcılık Sınırı	“Bazı yaratıcı çözümlerimi anlamamış”	3/10
		Anında Motivasyon	“Zaman geçmeden hemen sonuç alabilmek beni gerçekten motive eder”	9/10
		İlerleme Takibi	“İlerlediğimi sürekli görebilmek çok motive edici”	8/10
		Oyunlaştırma Etkisi	“Yazılı puanımı hemen görebilmek çok motive edici oyun gibi hissediyorum”	6/10
	Anlaşılma Zorluğu	Duygusal Değer	“Öğretmenimden övgü almak daha duygusal”	7/10
		Teknik Dil Sorunu	“Çok teknik açıklamalar yapıyor”	8/10
		Bilgi Yoğunluğu	“Çok fazla bilgi veriyor neyin önemli olduğunu anlayamıyorum”	7/10
		Karışıklık	“Çok detaylı ama bazen karışık bence”	6/10
ORTA	Esneklik İhtiyacı	Öncelik Belirsizliği	“Hangi kısmının önemli olduğunu ayırt edemiyorum”	5/10
		Katı Değerlendirme	“Çok katı değerlendirme yapıyor küçük hatalarımı büyütüyor”	7/10
		Hata Toleransızlığı	“İşlem hatası yaptığımda tüm soruyu yanlış sayıyor”	6/10
		Bireysel Fark Görmeme	“Bireysel farklılıkları görmüyor”	8/10
		Esnek Olmama	“Esnek değil farklı düşünce tarzlarını pek dikkate almıyor”	7/10

(Devam ediyor)

Tablo 13 (Devam)

Odak Grup Görüşmesi Geri Bildirim Analizi

GRUP	TEMA	KOD	ÖRNEK CÜMLE	FREKANS
ORTA	Motivasyon Karmaşası	Moral Bozucu Etkiler	<i>“Bazen aldığım düşük puanlar moralimi bozuyor”</i> <i>“Sürekli hata mesajları alınca motivasyonum düşüyor”</i>	6/10
		Cesaretsizlik	<i>“Çok pratik yapmama yardımcı oluyor ama başarısız olduğumda cesaretimi kırıyor”</i>	5/10
		Çelişkili Duygular	<i>“YZ karmaşık terimler kullanıyor kafam karışıyor”</i>	7/10
		Dil Karmaşıklığı Seviye Uyumsuzluğu	<i>“YZ çok zor kelimeler kullanıyor anlamıyorum”</i>	9/10
	Anlaşılmazlık	Basitlik Tercihi	<i>“YZ çok soğuk çok bilimsel”</i> <i>“Hiç anlayışlı değil”</i>	8/10
		Soğukluk Hissi	<i>“En ufak hatayı yanlış sayıyor bu hiç adil değil”</i>	7/10
		Anlayışsızlık	<i>“Bizi tanımadan nasıl eşit davransın”</i>	9/10
		Hata Büyütme	<i>“Öğretmenim benim anlayabileceğim şekilde açıklıyor”</i>	10/10
	Adaletsizlik Algısı	Tanınamama	<i>“Öğretmenim çabamı görüyor ona göre not veriyor”</i>	8/10
		Çaba Görmeme	<i>“YZ değerlendirmesinde başarısız hissediyorum”</i>	8/10
ALT	Motivasyon Kaybı	Başarısızlık Hissi	<i>“Verdiği puanlar da çok düşük güvenmiyorum”</i>	6/10
		Güvensizlik	<i>“Beni motive etmiyor”</i>	7/10
		Motivasyon Eksikliği	<i>“Öğretmenimin değerlendirmelerinde daha iyi hissediyorum”</i>	9/10
		Öğretmen Tercihi		

Öğrencilerin yanıtlarında sık tekrar eden temalar kodlanarak elde edilen bulgular YZ sistemlerinden alınan geri bildirimlerin öğrenci başarı düzeylerine göre belirgin farklılıklar gösterdiğini ortaya koymaktadır. Bu farklılıklar, teknolojinin eğitime entegrasyonunda yapılması gerekenler için yol göstericidir. Tek bir yaklaşımın tüm eğitim sisteminde işlevsel olmayacağı, eğitimdeki tüm araçlar gibi YZ sistemlerinin de öğrencilerin ihtiyaç ve seviyelerine göre farklılaştırılması gerektiği görülmektedir.

Üst düzey başarıya sahip olan gruptaki öğrenciler YZ sistemlerini fırsat olarak görerek sistemin hızlı, ayrıntılı ve tutarlı olmasını yüksek oranda değerli görmektedir (%80-90 kabul oranı). "YZ öğretmenden daha hızlı ve ayrıntılı geri bildirim verebilir" ifadesi sistemin verimliliğini; "hatalarımı adım adım gösteriyor" ifadesi öğrenme sürecinde YZ sistemlerini destekleyici bulduklarını belirtmektedir. Üst düzey başarıya sahip olan gruptaki öğrenciler, YZ'nin objektif ve önyargısız olduğunu; YZ geri bildirimlerinin ise öğretmen değerlendirmelerinden daha adaletli olduğunu düşünmektedir. Başarısı yüksek olan öğrencilerin YZ sistemlerinin hızlı olmasını motivasyon açısından olumlu buldukları gözlemlenmektedir.

Orta düzey başarıya sahip olan gruptaki öğrenciler, YZ sistemlerinin bir potansiyeli olduğunu fark etmekte ancak uygulama düzeyinde zorluklar yaşadıklarını belirtmektedir. "Çok teknik açıklamalar yapıyor" ve "neyin önemli olduğunu anlayamıyorum" ifadeleri, teknik dil kullanımının orta düzey başarıya sahip olan gruptaki öğrenciler için aşılması gereken bir durum olduğu görülmektedir. Geri bildirim verilirken öğrencinin seviyesine uygun dilin kullanılması önemlidir. Orta düzey başarıya sahip olan grubun YZ sistemleri ile ilgili adalet algısında bir ikilem olduğu görülmektedir. Öğrenciler hem YZ sisteminin objektif olduğunu düşünmekte hem de sistemi "çok katı" ve "esnek olmayan" olarak değerlendirmektedir. Başarısızlıklarını YZ sisteminin katılığına bağlamaları nedeni ile sistemi derslerde pratik yapmak için kullanabileceklerini ancak puanlama ve değerlendirmelerde tercih etmeyeceklerini belirtmişlerdir.

Alt düzey başarıya sahip olan grup öğrencilerin, YZ sistemlerine yönelik olumsuz algıları olduğu gözlemlenmiştir. "YZ karmaşık terimler kullanıyor, kafam karışıyor" ifadesi, sistemin bu grup için uygun olmadığını gösteren yanıtlardandır. Alt grubun adalet algısı, "hiç anlayışlı değil" ve "en ufak hatayı yanlış sayıyor" gibi ifadelerle bakılarak yorumlanırsa, öğrencilerin "adil" olmayı diğer gruplardan farklı gördüğü düşünülebilir. Sistemi anlayışsız bulmaları, kendi çaba ve gayretlerinin görülmediği duygusu öğrencilerin motivasyonunu düşürmekte ve başarısızlık hissi yaratarak öğrenme sürecine zarar vermektedir.

Üst başarı grubundaki öğrenciler, YZ'nin ayrıntılı geri bildirim verdiğini belirtmiş, geri bildirim kalitesine ilişkin genellikle olumlu yönde görüşler bildirmişlerdir. Alan yazında yapılan bir çalışmada da benzer bir bulguya ulaşılmıştır. Baral ve diğerleri (2025), çalışmalarında, matematik eğitiminde açık uçlu sorulara verilen öğrenci yanıtları için otomatik geri bildirim üretmede LLM'lerin etkinliğini araştırmak amacıyla üç farklı modeli karşılaştırmıştır (SBERT-Canberra, GOAT ve GPT-4). Geri bildirim kalitesinde

GPT-4'ün açık ara önde olduğu ortaya konmuştur. Öğretmenlerin bu sistemlerin sağladığı geri bildirimlere yönelik olumlu değerlendirmeleri dikkat çekmektedir. Araştırmada öğretmenler, özellikle gelişmiş dil modellerinin ürettiği geri bildirimlerin kapsamlı ve derinlemesine olduğunu, öğrenci çalışmalarının güçlü ve zayıf yönlerine dair değerli içgörüler sunduğunu ifade etmişlerdir. Çalışmaya katılan öğretmenler, YZ destekli geri bildirimi "etkili bir başlangıç noktası" olarak tanımlarken, bu sistemin geri bildirim oluşturma sürecindeki zihinsel yükü önemli ölçüde azalttığını belirtmiş; YZ sistemlerinin ürettiği geri bildirimlerin genel olarak doğru, ilgili ve öğrenci motivasyonunu destekleyici nitelikte olduğunu vurgulamıştır.

Ancak tüm görüşme sorularında verilen yanıtlarda dikkat çekici olan, tüm grup öğrencilerin öğretmeni tamamen değiştirilemez olarak görmesidir. Bu durum, eğitimde YZ sistemlerinin "öğretmenin yerini alması" değil, "öğretmene yardımcı" rolünde konumlandırılması gerektiğini göstermektedir. Alt grup öğrencilerinin duygusal destek ihtiyacı, teknolojik çözümlerin günümüzde hala insanın yerine geçemeyeceğini ortaya koymaktadır. YZ sistemlerinin eğitime başarılı entegrasyonu, öğrenci seviyelerine göre farklılaştırılmış tasarım, içerik ve sunum yaklaşımları gerektirmektedir. Bu sistem, öğrenci başarı düzeyi yükseldikçe daha özerk, düştükçe daha destekleyici ve rehberlik odaklı olarak tasarlanmalıdır.

YZ Tabanlı Değerlendirme Sistemlerinin Eğitim Süreçlerine Entegrasyonu

Bu bölümde, YZ tabanlı değerlendirme sistemlerinin günümüzde eğitim süreçlerine eklemlenmesi süreci ve bu süreçte yaşanan zorluklar incelenmiştir. Bu sürecin öğretmenler ve öğrenciler tarafından algılanışı ile uygulamada ortaya çıkan teknik ve pedagojik zorlukların neler olduğu analiz edilmiştir. Çalışmada, öğretmen görüş anketi, öğretmenlerle derinlemesine görüşme ve öğrenciler ile odak grup çalışması yapılarak veriler toplanmıştır. Elde edilen bulgular, YZ tabanlı değerlendirme sistemlerinin eğitim ekosistemine uyumlu bir şekilde dahil edilmesi için gerekli koşulları ve stratejileri ortaya koymaktadır.

Araştırmanın üçüncü alt amacı olan “YZ tabanlı değerlendirme sistemleri eğitim ortamlarında hangi yollarla uygulanmakta ve hangi amaçlarla kullanılmaktadır?” sorusu, Türkiye’nin farklı şehirlerinde görev yapan farklı branşlardan öğretmenlere çevrimiçi form aracılığı ile ulaştırılarak anket sonuçları alınmıştır. 4’lü likert tipi ölçekte “1. Hiç Katılmıyorum, 2. Katılmıyorum, 3. Katılıyorum, 4. Tamamen Katılıyorum” şeklinde

derecelendirilerek elde dilen anket verileri madde frekansları, yüzdeleri, ortanca ve çeyrek açıklık bilgileri Tablo 14’te sunulmuştur.

Tablo 14

Öğretmen Anketinde Yer Alan Her Maddeye İlişkin Verilerin Yanıtların Dağılımına İlişkin Frekans ve Yüzdeler

Madde	1 f(%)	2 f(%)	3 f(%)	4 f(%)	Ortanca	Q1	Q3
1. YZ tabanlı puanlama sistemlerinin mevcut eğitim süreçlerine entegrasyonu başarılıdır.	1 (2.8)	7 (19.4)	20 (55.6)	8 (22.2)	3	3.00	3.00
2. YZ destekli puanlama sistemlerinin kullanılması öğretim ve değerlendirme süreçlerini iyileştirir.	0 (0)	4 (11.1)	17 (47.2)	15 (41.7)	3	3.00	4.00
3. YZ ile yapılan değerlendirmeler öğretmen değerlendirmesi kadar iyi değildir.	7 (19.4)	13 (36.1)	13 (36.1)	3 (8.3)	2	2.00	3.00
4. YZ destekli puanlama sistemleri eğitimcilerin iş yükünü azaltır.	1 (2.8)	2 (5.6)	12 (33.3)	21 (58.3)	4	3.00	4.00
5. YZ destekli puanlama sistemleri her okulda kullanılamayacağı için eğitim alanına entegre edilmemelidir.	0 (0)	12 (33.3)	11 (30.6)	13 (36.1)	3	2.00	4.00
6. YZ destekli sistemler öğrencilere etkili geri bildirimler sağlar.	0 (0)	4 (11.1)	18 (50)	14 (38.9)	3	3.00	4.00
7. YZ tabanlı değerlendirme sistemlerinin kullanımı zordur.	2 (5.6)	5 (13.9)	12 (33.3)	17 (47.2)	3	3.00	4.00
8. YZ destekli puanlama sistemlerinin eğitimde daha etkin kullanılabilmesi için teknik ve pedagojik iyileştirmeler gereklidir.	0 (0)	3 (8.3)	9 (25)	24 (66.7)	4	3.00	4.00
9. YZ destekli puanlama sistemleri eğitimde kullanılmalıdır.	0 (0)	9 (25)	12 (33.3)	15 (41.7)	3	2.25	4.00
10. Bu teknolojilerin eğitim alanında geniş çapta kabul görmesi gereklidir.	0 (0)	5 (13.9)	9 (25)	22 (61.1)	4	3.00	4.00

Araştırmada, öğretmenlerin branşlarına göre YZ tabanlı değerlendirme sistemlerine karşı tutumlarının belirlenmesi amacıyla uygulanan ankete toplam 36 öğretmen katılmıştır. Betimsel istatistikler sonucunda öğretmenlerin YZ sistemlerinin eğitimde kullanımına yönelik genel tutumlarının olumlu olduğu görülmektedir. Likert ölçeğinin ordinli (sıralı) yapısı nedeniyle merkezi eğilim ölçüsü olarak medyan değerleri kullanılmıştır. Anket maddelerinin tek tek değerlendirilmesi ile öğretmenler arasında en yüksek olumlu tutumun 8. maddede (Pedagojik iyileştirmeler gerekli) görüldüğü tespit edilmiştir (medyan = 4, Q1 = 3.00, Q3 = 4.00). Benzer şekilde, 4. madde (İş yükünü azaltır, medyan = 4, Q1 = 3.00, Q3 = 4.00) ve 10. madde (Geniş çapta kabul görmeli, medyan = 4, Q1 = 3.00, Q3 = 4.00) de tüm branşlarda YZ ile ilgili güçlü olumlu görüşleri yansıtmaktadır. Öğretmenler arasında en olumsuz tutum ise 3. maddede (YZ öğretmen kadar iyi değildir, medyan = 2, Q1 = 2.00, Q3 = 3.00) görülmüştür. Öğretmenler, YZ değerlendirmelerini öğretmen değerlendirmeleri kadar etkili görmemektedir.

En homojen görüşlerin 1. maddede (Entegrasyon başarılı, Q1 = 3.00, Q3 = 3.00, IQR = 0.00) olduğu, yani bu konuda öğretmenlerin tam fikir birliği içinde olduğu tespit edilmiştir. En heterojen görüşler ise 5. maddede (Her okulda kullanılamaz, Q1 = 2.00, Q3 = 4.00, IQR = 2.00) ve 9. maddede (Eğitimde kullanılmalı, Q1 = 2.25, Q3 = 4.00, IQR = 1.75) görülmüştür. Bu maddeler öğretmenler arasında en çok tartışılan konuları yansıtmaktadır.

Gruplanan branşlar açısından sıra ortalaması analizi yapıldığında, Ölçme ve Değerlendirme branşının en iyi performansı gösterdiği (sıra ortalaması: 4.00), bunu Bilişim Teknolojileri (sıra ortalaması: 7.00) ve Türkçe (sıra ortalaması: 12.50) branşlarının izlediği tespit edilmiştir. Matematik branşı en geniş performans yelpazesine sahip olup (1. sıradan 36. sıraya), bu alanda görüş farklılığının en yüksek olduğu branş olmuştur. Martínez-Comesaña ve diğerlerinin (2023) sistematik literatür taramasında, YZ'nin ilk ve ortaöğretim seviyelerinde öğrenci değerlendirmesinin iyileştirilmesine yönelik çok disiplinli bir alan oluşturduğu ve eğitim sektörünün YZ'nin çeşitli yönlerde dahil edilmesiyle zenginleştirilebileceği vurgulanmıştır. Bu bulgular, öğretmenlerin YZ sistemlerine yönelik genel olarak olumlu tutum sergilediği ve özellikle pedagojik iyileştirmeler ve iş yükü azaltma konularında bu sistemleri desteklediği yönündeki bulguları ile tutarlılık göstermektedir.

Çeyreklikler arası açıklık (IQR) analizleri, öğretmenlerin hangi konularda uzlaşısı içinde olduğunu, hangi konularda görüş ayrılığı yaşadığını net şekilde ortaya koymuştur. IQR değerleri 0.00 ile 2.00 arasında değişim göstermiş, bu da bazı konularda tam fikir

birliğı (IQR = 0.00) olduğunu, bazı konularda ise önemli görüş farklılıkları (IQR = 2.00) bulunduğunu göstermektedir. En güçlü anlaşmanın "Entegrasyon başarılı" ve "Pedagojik iyileştirmeler gereklidir" maddelerinde olduğu görülmektedir. Branş bazında sıra ortalaması analizinde ise branşlar arası performans farkları tespit edilmiş, ancak bu farkların sistematik bir desen izlemediği gözlemlenmiştir.

Sonuç olarak, bu anket sonucu öğretmen branşlarının YZ sistemlerine yönelik tutumlarında istatistiksel olarak anlamlı farklılık yaratmadığını ortaya koymuştur. Tüm branş grupları benzer şekilde olumlu tutum sergilemekte, pedagojik iyileştirmeler gerektiği konusunda güçlü anlaşma göstermekte ve öğrenci odak grup görüşmeleri ile paralel şekilde YZ sistemlerinin öğretmenleri destekleyici rolde olması gerektiğini düşünmektedir. Bu bulgular, YZ tabanlı puanlama sistemlerinin eğitime entegrasyonunda branş ayrımı yapmayan homojen stratejilerin benimsenebileceğini, ancak öğretmen-YZ işbirliği modelinin güçlendirilmesi gerektiğini işaret etmektedir.

Dördüncü alt amaç olan “Öğretmenler ve öğrenciler, YZ tabanlı değerlendirme sistemlerinin eğitim süreçlerine entegrasyonunu nasıl değerlendirmektedir?” sorusu için öğrencilerle yapılan odak grup görüşmesine ilişkin verdikleri yanıtların nitel veri analizi Tablo 15’te sunulmuştur.

Tablo 15

Öğrencilerle Yapılan Odak Grup Görüşmesi Entegrasyon Süreçleri Analizi				
Öğrenci başarı düzeyi	Tema	Kod	Örnek Cümle	Frekans
ÜST	Hibrit model	Tamamlayıcılık	“İkisi birbirlerini tamamlıyorlar”	10/10
		Fonksiyonel ayırım	“YZ günlük çalışmalar için öğretmen genel değerlendirme için”	8/10
		Dengeli kullanım	“İkisini dengeli kullanmayı tercih ederim”	7/10
		Hız-kalite kombinasyonu	“Hem hız hem de kalite bir arada olur”	6/10
	Özgürlük kontrol	Yargısızlık rahatlığı	“Bir yazılım ve beni yargılamaz. Hata yapmaktan korkmuyorum”	10/10
		Kişisel tempo	“Kendi tarzıma göre çalışabilirim”	9/10
(Devam ediyor)				

Tablo 15 (Devam)

Öğrencilerle Yapılan Odak Grup Görüşmesi Entegrasyon Süreçleri Analizi

Öğrenci başarı düzeyi	Tema	Kod	Örnek Cümle	Frekans
ÜST	Özgürlük Kontrol	Tekrar özgürlüğü	“Aynı soruyu defalarca sorabilirim ”	8/10
		Sosyal baskısızlık	“Sosyal baskı hissetmiyorum ”	7/10
		Görsel takip	“Her şeyi görselleştirebiliyor grafik gibi ”	9/10
		Otomatik kayıt	“Her şey otomatik kaydediliyor ”	10/10
	İlerleme takip	Konu bazlı analiz	“Hangi konularda zayıf olduğumun net bir şekli olur ”	8/10
		Geçmiş erişimi	“Geçmişe dönüp baştan bakabilirim ”	7/10
		Teknoloji kabulü	“Geleceğin olayı YZ ”	8/10
		Öğretmen varlığı	“Öğretmenin yerini almamalı eğitimde insan olmadan olmaz ”	10/10
	Gelecek vizyonu	Verimlilik artışı	“Daha verimli bence ”	9/10
		Sistem entegrasyonu	“Mükemmel bir eğitim sistemi olur ”	7/10
		Öğretmen önceliği	“Asıl değerlendirmeyi öğretmenim yapsın, YZ sistemi öğretmenimin yardımcısı olsun ”	8/10
		YZ yardımcı rol	“YZ sistemi yardımcı olarak iyi ”	7/10
ORTA	Öğretmen ağırlıklı model	Güven faktörü	“Güven açısından daha iyi ”	6/10
		Rahatlatıcı etki	“Bu daha rahatlatıcı ”	5/10
		Zaman özgürlüğü	“Zaman açısından özgürüm ”	9/10
		Yöntem kısıtı	“Çözüm yöntemlerinde kısıtlıyım ”	8/10
	Kısıtlı özgürlük	Yaratıcılık engeli	“Yaratıcı çözümlerimi anlamıyor ”	7/10
		Kalıp zorlaması	“Beni kalıplara sokuyor ”	6/10
	Rehberlik ihtiyacı	Veri yorumlama zorluğu	“Bu verileri nasıl yorumlayacağımı bilmiyorum ”	8/10

(Devam ediyor)

Tablo 15 (Devam)

Öğrencilerle Yapılan Odak Grup Görüşmesi Entegrasyon Süreçleri Analizi

Öğrenci başarı düzeyi	Tema	Kod	Örnek Cümle	Frekans
ORTA	Rehberlik ihtiyacı	Çözüm önerisi eksikliği	“Sadece problemi söylüyor çözüm sunmuyor”	7/10
		Gelişim stratejisi yokluğu	“Gelişim stratejileri konusunda rehberlik etmiyor”	6/10
		Geliştirme şartı	“Geliştirilirse daha fazla kullanabilirim”	8/10
	Şartlı kabul	Basitleştirme isteği	“Daha kullanıcı dostu olmalı”	7/10
		İnsan dokunuşu	“İnsan dokunuşu eksik”	6/10
		Soğukluk eleştirisi	“Çok robotik geliyor”	5/10
	Öğretmen odaklı	İnsan ihtiyacı	“Öğretmenimle yüz yüze olmayı isterim destekleyici oluyor”	9/10
		Duygusal bağ	“Öğretmenim beni daha iyi tanıyor”	8/10
		Sıcaklık arayışı	“YZ bana soğuk geliyor”	7/10
		Güven duygusu	“Öğretmenime güveniyorum”	8/10
ALT	Karma özgürlük	Utanma özgürlüğü	“O bir makine ve utanmama gerek yok hata yapınca”	4/10
		Kısıtlanma hissi	“Kısıtlanmış hissediyorum”	6/10
		Baskı algısı	“Sürekli kontrol altında gibi hisseder insan”	5/10
	Olumsuz takip	Kararsızlık	“Emin değilim”	3/10
		Moral bozucu sonuçlar	“Kötü sonuçlar aldığım zaman bunların sürekli karşımda olması beni mutlu etmez”	8/10
		İlerleme yokluğu	“İlerleme kaydetmiyorum ki hep aynı yerde kalıyorum”	7/10
		Takip istememe	“Takip etmek istemiyorum çünkü hep kötü sonuçlar görüyorum”	6/10
		Öğretmen üstünlüğü	“Öğretmenim YZ'den daha iyi bu konuda cesaret de veriyor”	9/10

(Devam ediyor)

Tablo 15 (Devam)

Öğrencilerle Yapılan Odak Grup Görüşmesi Entegrasyon Süreçleri Analizi

Öğrenci başarı düzeyi	Tema	Kod	Örnek Cümle	Frekans
ALT	Düşük kabul	Gerek görmeme	"Bence gerek yok Öğretmen değerlendirmesi daha iyi"	6/10
		Seviye uyumu isteği	"Bizim seviyemize göre olmalı"	4/10
		Basitleştirme şartı	"Basitleştirse belki kullanabilirim"	3/10
		Gelecek umudu	"Gelecekte daha öğrenciye kolay gelecek bir hale gelirse neden olmasın"	2/10

Odak grup görüşmesi sonuçları, başarı seviyesi gruplarına göre bazı farklılıklar göstermektedir. Üst düzey başarıya sahip gruptaki öğrenciler, hibrit modeli güçlü bir şekilde desteklemekte ve "ikisi birbirlerini tamamlıyorlar" şeklinde görüşleri ile her iki değerlendirmenin de farklı avantajlarından faydalanmak istemektedirler. Bu bulgu alan yazında yapılan çalışmaların bulgularıyla tutarlılık göstermektedir. McNichols vd., (2024) LLM'lerin akıllı öğretim sistemlerinde (ITS) kullanılan önceden hazırlanmış mesaj metinlerini kullanarak öğrencilere otomatik geri bildirim sağlama sürecini taklit edebilme kapasitesi üzerine yaptıkları çalışmada, LLM'lerin insanları destekleyecek araçlar olarak kullanılması gerekliliği sonucuna varılmıştır ancak mevcut çalışma öğrenci perspektifinden hibrit modele yönelik tutumları inceleyen ilk çalışmalardan biridir.

Üst düzey başarıya sahip gruptaki öğrenciler, YZ sistemlerinin sağladığı "yargısızlık rahatlığı" ile kendilerini daha özgür hissettiklerini belirtmekte; hata yapma, utanma korkusu yaşamadıklarını ve kendi hızlarında, kendilerine özgü çalışabildiklerini vurgulamaktadırlar. Ayrıca üst düzey başarıya sahip gruptaki öğrenciler, verilerin sürekli kayıt edilip ulaşılabilir olması ve verilerin görselleştirilerek grafiklerle desteklenmesi gibi unsurların öğrenme süreçlerini ilerlettiğini ve ilerleyen zamanda YZ entegrasyonunun kaçınılmaz olduğunu kabul etmektedirler.

Orta düzey akademik başarı göstermiş olan öğrenciler sürece daha temkinli yaklaşmaktadır. YZ sistemlerini sürecin dışında bırakmayı istememekle birlikte öğretmen ağırlıklı bir modeli seçerek "asıl değerlendirmeyi öğretmenim yapsın, YZ sistemi öğretmenimin yardımcısı olsun" görüşünü savunmaktadırlar. Orta düzey başarıya sahip

gruptaki öğrenciler, YZ sistemlerinin hızlı olmasından dolayı zaman özgürlüğü sağladığını kabul etseler de farklı çözümlerinin anlaşılmadığını ve sistemin belli kalıplara göre değerlendirdiğini düşünmektedirler. YZ sisteminin geri bildirimlerinde yer alan verilerin karmaşık olduğunu ve bu verileri yorumlamakta zorlandıklarını belirten orta düzey başarıya sahip gruptaki öğrenciler bu sistemden yeterli rehberlik alamadıklarını belirtmiş, sistemlerin geliştirilmesi halinde daha fazla kullanabileceklerini söylemişlerdir.

Alt düzey başarıya sahip gruptaki öğrenciler ise YZ sistemlerine karşı dirençli bir tavır sergilemektedir. Bu öğrenciler, "öğretmenimle yüz yüze olmayı isterim, destekleyici oluyor" düşüncesi nedeniyle bir insanla karşı karşıya olma ihtiyacını ön plana çıkararak öğretmenleriyle kurdukları duygusal bağın önemini vurgulamaktadırlar. Buna bağlı olarak YZ sistemlerini "soğuk" ve "robotik" olarak nitelendirmiş; sürekli kontrol altında hissettiklerini ve utanmalarına gerek olmayan bir ortam olduğunu kabul etseler de kısıtlanma duygusu yaşadıklarını belirtmişlerdir. Bununla birlikte kötü sonuçların sürekli karşılarında olmasının motivasyonlarını düşürdüğünü ve bu nedenle takip etmek istemediklerini ifade etmişlerdir.

Bu bulgular, akademik başarı düzeyi ile YZ tabanlı değerlendirme sistemlerine karşı tutum arasında güçlü bir ilişki olduğu yönünde yorumlanabilir. Üst başarı düzeyindeki öğrencilerin teknoloji ile uyum sağlama kapasitelerinin yüksek olduğu, YZ sistemlerinin sunduğu tarafsız değerlendirmeyi olumsuz olarak algılamadıkları anlaşılmaktadır. Orta başarı düzeyindeki öğrencilerin kendi yararlarına olan kısımları kabullenmeleri ancak temkinli davranarak hem teknolojinin avantajlarını reddetmedikleri hem de güvenlik ihtiyacı hissettiklerini göstermektedir. Alt başarı düzeyindeki öğrencilerin karşı çıkması ise YZ sistemlerinin tarafsız değerlendirmelerinin bu öğrencilerde kaygı yarattığını ve öğretmenlerinden aldıkları duygusal desteğe daha fazla ihtiyaç duyduklarını göstermektedir. Bu bulgular, YZ sistemlerinin farklı akademik seviyelerdeki öğrencilerin istek ve gerekliliklerine göre farklılaştırılması ve özellikle alt grup öğrenciler için destekleyici mekanizmalarının geliştirilmesi gerektiğini ortaya koymaktadır.

Dördüncü alt amaç olan “Öğretmenler YZ tabanlı değerlendirme sistemlerinin eğitim süreçlerine entegrasyonunu nasıl değerlendirmektedir?” ve beşinci alt amaç olan “YZ tabanlı değerlendirme sistemlerinin entegrasyonu sürecinde hangi teknik ve pedagojik zorluklarla karşılaşılacaktır?” soruları için yazılı sınavın uygulandığı özel okulun ölçme uzmanları (U) ve matematik öğretmenleri (Ö) ile birebir görüşme yapılmış, görüşme sorularına verilen yanıtların nitel analiz tablosu Tablo 16’da sunulmuştur.

Tablo 16

Ölçme Uzmanları ve Matematik Öğretmenleri İle Yapılan Birebir Görüşmenin Analizi

Tema	Alt kod	Örnek Alıntı
YZ deneyimi	Kullanmamış	“Hayır hiç kullanmadım. Geleneksel yöntemlerle değerlendirme yapıyorum” (Ö1)
	İstekli	“Çok merak ediyorum deneme fırsatı arıyorum” (Ö4)
	Tecrübeli	“Evet kullandım. Değerlendirme konusunda oldukça yararlı bir sistem” (U1)
	Araştırma halinde	“Hayır kullanmadım ama aktif olarak araştırıyorum” (U2)
Eğitim sistemindeki yeri	Destekleyici rol	“Destekleyici bir araç olarak kullanılabilir ama ana değerlendirme yöntemi olamaz” (Ö1)
	Gelecekte önemli	“Bence gelecekte çok önemli olacak. Eğitimin dijitalleşmesi kaçınılmaz” (Ö2)
	Altyapı eksiklikleri	“Türkiye’de ciddi altyapı eksiklikleri var. Öğretmenlerimiz hazır mı emin değilim” (U2)
Başarı ölçmede yeterlilik	Süreç - sonuç	“Matematik sadece sonuç değil süreç de çok önemli. YZ bunları görebilir mi emin değilim” (Ö1)
	Basit - karmaşık	“Basit standart sorularda belki evet ama karmaşık problemlerde şüpheliyim” (Ö3)
	Mantık - düşünce	“Matematik mantık ister düşünme becerisi ister. YZ gerçekten mantığı anlayabilir mi” (Ö5)
	Adalet - eşitlik	“Adalet değil eşitlik sağlama sorun olabilir. Sınıf içi durumu öğrencinin kaygılarını bilemez” (U1)
Değerlendirme karşılaştırması	İnsan faktörü	“Öğretmenin kalbi var YZ sadece kodlardan oluşuyor. İnsan faktörü değiştirilemez” (Ö5)
	Empati	“Ben öğrencimin psikolojisini biliyorum o günkü halini anlıyorum. YZ bunu bilemez” (Ö3)
	Objektiflik	“YZ daha objektif olabilir, öğretmen bazen duygusal yaklaşabilir” (Ö1)
	Hevesli	“Kesinlikle entegre etmek isterim! Yenilikçi olmak istiyorum” (Ö2)
Eğitime entegrasyon	Temkinli	“Biraz zorluyor beni. Teknoloji öğrenmek bu yaşta kolay değil” (Ö3)
	İş yükü hafifletme	“Benim işimle ilgili olarak iş yükümü kesinlikle hafifletir” (U1)
	Eğitim gerekliliği	“Önce kapsamlı öğretmen eğitimi şart. Sistem hazır değilse büyük sorunlar çıkar” (U2)

(Devam ediyor)

Tablo 16 (Devam)

<i>Ölçme Uzmanları ve Matematik Öğretmenleri İle Yapılan Birebir Görüşmenin Analizi</i>		
Tema	Alt kod	Örnek Alıntı
Geri Bildirim ve Motivasyon	Anında geri bildirim	“Anında geri bildirim çok büyük avantaj. Öğrenci hemen öğrenir hatasını” (Ö2)
	Kişiselleştirme eksikliği	“Ben her öğrencimle özel konuşurum onun seviyesine göre açıklama yaparım” (Ö1)
	Duygusal risk	“Başarı seviyesi düşük öğrenciyi kırabilir çünkü duygusal bir yanı yok” (U1)
Pedagojik uyumluluk	Bireysel farklılıklar	“Her öğrenci farklı öğrenir. Kimisi görsel kimisi işitsel. YZ bu farklılıkları anlayabilir mi” (Ö1)
	Yaşanmış deneyim	“Pedagoji teorik bilgi ile olmaz, yaşanmışlık ve deneyimle olur. YZ bu deneyimi yaşayamayacağı için sorun olabilir” (U2)
	Kültürel faktörler	“Türk öğrencisinin özelliklerini, kültürümüzü öğrenebilir mi bilmiyorum. Pedagoji bunlara da bağlı” (Ö3)
Sistem geliştirme önerileri	Teknik özellikler	“Türkçe olmalı kesinlikle. Adım adım çözüm göstermeli. Matematik işlem adımlarını gösterebilmeli” (Ö1)
	Hibrit model	“Mutlaka öğretmen kontrolünde olmalı, öğretmenlerin yönergeleri ile gelişir” (U3)
	Öğretmen eğitimi	“Öğretmen eğitimi mutlaka verilmeli. YZ’nin bizi öğrenmesi de gerekli bizim onu öğrenmemiz de” (Ö5)
Motivasyon	Yüksek motivasyon	“Yenilik beni heyecanlandırıyor. Gelecek bu yönde” (Ö4)
	Karışık duygular	“Bazen kolaylaştırır bazen işleri karmaşık hale getirir diye düşünüyorum” (Ö3)
	Mesleki kaygı	“İş açısından kolaylık olur ama mesleki tatmin azalır mı bilmiyorum” (U3)
	Pragmatik kabul	“Ben zaten çok çalışıyorum benim için motive edici ve gerekli” (U1)

Yapılan görüşmelerde öğretmenlerin, öğrenci-öğretmen ilişkisinde duygusal ve psikolojik boyuta vurgu yaptıkları ve öğretimin bütünlüğünün korunması konusunda kaygı yaşadıkları görülmektedir (*Matematik eğitiminde öğretmenin rolü değiştirilemez. - Öğretmenin kalbi var, YZ kodlardan oluşuyor*). Ölçme uzmanları ise verilerin doğru işlenmesi, analizlerin kaliteli olması ve ölçme sonuçlarının değerlendirilmesi süreçlerinde tarafsızlığın ve netliğin olması konularını önemsemekle birlikte duygusal tatmin açısından öğretmenlerle aynı kaygıyı yaşamaktadır.

YZ sistemlerine yaklaşım konusunda öğretmenler arasında kuşak farkları gözlemlenmektedir. Katılımcı öğretmenlerden yaşı daha genç olanlar bu sistemlerin

eğitime entegre edilmesi konusunda olumlu görüşler sergilerken deneyimli öğretmenler daha temkinli yaklaşarak insanların eğitim alanında vazgeçilmez olduğunu özellikle vurgulamaktadır. Pedagojinin önemli olmadığı veri işleme ve iş yükünün azalması gibi konular söz konusu olduğunda ölçme uzmanları pragmatist bir yaklaşım sergileyerek YZ sistemlerinin yararlarını ön plana almaktadır.

Öğretmenlerin de ölçme uzmanlarının da öğrenci ile kurulan ilişkiler konusunda görüşleri benzeşmektedir. Öğretmenler öğrenciyi tanımak ve çabasına göre, psikolojisine göre yaklaşmak gerektiğini ancak YZ sistemlerinin bunu yapamayacağını düşünmektedir (*Ben öğrencimin psikolojisini biliyorum, o günkü halini anlıyorum. - Ben her öğrencimle özel konuşurum, onun seviyesine göre açıklama yaparım*). Aynı şekilde ölçme uzmanları da sistemlerin tarafsız olmasının sorun olabileceği durumlar ile ilgili kaygılarını belirtmektedir (*adalet değil eşitlik sağlama sorun olabilir*).

YZ sistemlerinin iş yükünü hafifletmesi herkes tarafından kabul edilmekle birlikte öğretmenler YZ'nin destek kuvvet olarak işe yarayacağını, öğretmenin yerini alamayacağını özellikle belirtmektedir. Teknolojik araçlarla sürekli temas halinde olmak hem meslek hem de yaş açısından farklılık göstermektedir. Öğretmenlerin YZ sistemlerine ihtiyaç olmadan işlerini yürütme görüşü ölçme uzmanlarına göre daha baskın görünmektedir. Çoğunlukla teknolojik araçlar kullanarak işlerini sürdüren ölçme uzmanları ise yeni bir teknoloji ile iş yüklerinin hafiflemesi fikrine daha olumlu yaklaşmaktadır. Görüşme sorularına verilen yanıtlar doğrultusunda öğretmenlerin YZ sistemlerini kabullenip derslerine entegre etmeleri, ölçme uzmanlarının işlerinde bu sistemleri kullanmasından daha uzun bir süreç gibi görünmektedir.

BÖLÜM 5

SONUÇ VE ÖNERİLER

Bu bölümde, araştırmadan elde edilen bulgulara dayalı sonuçlara yer verilmiş ve bu sonuçlara dayanarak öneriler sunulmuştur.

Sonuçlar

Bu araştırmada, açık uçlu soruların değerlendirilmesi ve geri bildirim verilmesi sürecinde yapılan analizler ile elde edilen bulgular, YZ teknolojisinin eğitimde kullanılması konusunda yol gösterici olması açısından önemlidir. Başarı testlerinde elde edilen sonuçların güvenilirliği, eğitim kararları alınırken bu testlerin sonuçlarının kullanılabilmesi için gerekli bir bilgidir. Matematik dersi 7. sınıf yazılısının değerlendirilmesi sonucunda YZ destekli sistemin (*Calude Sonnet*) güvenilirlik katsayısı, öğretmen değerlendirmesinin güvenilirlik katsayısından daha yüksek bulunmuştur. Cronbach'ın alfa katsayısı iki sistem için de "iyi" düzeyde güvenilirlik göstermektedir. Bu durum; açık uçlu sorularda, öğrencilerin kendi el yazıları ile yanıtladıkları yazılı belgelerinin puanlanmasında, YZ sistemlerinin tutarlı puanlama yapabilme becerisinin öğretmen değerlendirmelerine yakın olduğunu, hatta bazı durumlarda daha yüksek olabileceğini göstermektedir. Güvenirliğin daha yüksek olabileceği yorumu ise madde-toplam korelasyonları incelendiğinde görülmektedir. YZ puanlamasında çoğu maddede öğretmen değerlendirmesinden daha yüksek korelasyon göstermektedir. Bu durum, YZ sisteminin maddeleri testin geneli ile daha tutarlı bir şekilde değerlendirdiğini göstermektedir.

Testin güvenilirliği ve madde toplam korelasyonları yanında YZ sisteminin puanlamalarında madde ayırt edicilik gücünün de genel olarak öğretmen puanlamasından daha yüksek olduğu görülmüştür. Özellikle tarafsız bir değerlendirme gerektiren, çok aşamalı puanlama adımlarına sahip sorularda YZ sisteminin ayırt edicilik gücü belirgin şekilde yüksektir. YZ sistemin puanlamasında bu maddelerde daha yüksek ayırt edicilik görülmesi, tutarlı ve tarafsız puanlama yapabilme becerisi ile açıklanabilir. Ancak tüm maddelerin hem öğretmen hem de YZ puanlamasında "Çok İyi" ayırt edicilik düzeyinde

yer alması, her iki puanlayıcının da etkili bir şekilde üst ve alt grup öğrencilerini ayırt edebildiğini göstermektedir.

Öğrencilerin yanıtlarının değerlendirilmesinde, puanlayıcıların puanlarının uyumunu görmek amacı ile Ağırlıklandırılmış Kappa analizi yapılmıştır. Bu analiz sonucunda öğretmen ve YZ puanları arasındaki uyumun maddelere göre farklılaştığı görülmüştür. Madde analizlerinde toplam korelasyon ve ayırt edicilik indeksinin yüksek olması YZ puanlamasının tutarlı olduğunu göstermektedir. Bunun yanında, YZ sisteminin belirli adımları olan düzenli işlemlerde ve formüle dayalı sorularda öğretmenlerle yüksek uyum gösterdiğini ancak kavramsal anlama ve çok aşamalı problem çözme gerektiren sorularda farklı değerlendirmeler yaptığı tespit edilmiştir.

Araştırma bulgularına göre, YZ ve öğretmen puanlamaları arasındaki uyum düzeyinin konulara göre farklılık gösterdiği tespit edilmiştir. Birinci dereceden denklem kurma, yüzde hesaplamaları ve bir çokluğu yüzde ile arttırma/azaltma konularında her iki değerlendirme arasında güçlü düzeyde uyum varken; denklem çözme, doğru orantı, ters orantı, yüzde ile ilgili problemler ve doğru-ters orantı ile ilgili problem çözme konularında orta düzeyde uyum gözlemlenmiştir. Bununla birlikte, sözel ifadeden denklem kurma ve orantı hesaplaması gibi daha karmaşık matematiksel muhakeme gerektiren konularda puanlayıcılar arasında hafif düzeyde uyum olduğu belirlenmiştir. Bu durum, YZ sistemlerinin cebirsel işlemlerde öğretmen değerlendirmeleriyle daha yüksek tutarlılık gösterirken, sözel matematiksel ifadeleri yorumlama gerektiren konularda daha fazla farklılık sergilediğini ortaya koymaktadır. Analiz tablolarında da görüldüğü üzere test güvenilirliği, madde toplam korelasyonları ve madde ayırt ediciliklerinin öğretmen değerlendirmelerinden yüksek olması nedeniyle YZ sistemlerinin ön yargısız ve puanlama anahtarına uyarak puanlama yaptıkları düşünülebilir.

Bu araştırmanın bulguları, YZ sistemlerinin psikometrik açıdan öğretmen puanlamalarından daha yüksek bir puanlama performansı ortaya koyduğunu göstermektedir. YZ sisteminin güvenilirlik katsayısının yüksek olması, madde-toplam korelasyonlarının çoğunlukla daha yüksek değerler göstermesi ve ayırt edicilik gücünün genel olarak üstün olması, tutarlılık ve objektiflik açısından avantajlı konumda olduğunu göstermektedir. Ancak sayısal üstünlük ile değerlendirme kalitesinin sadece psikometrik göstergelerle ölçülüp ölçülemeyeceği gerçeği tartışılmalıdır. Bazı maddelerde yaşanan yüksek puan farkları, YZ sistemlerinin temel sınırlılıklarını göstermektedir. YZ sistemleri, sabit adımlar ve önceden belirlenmiş puanlama kriterleri ile çalışan sistemlerdir. Bu sistemler "Eğer X varsa → 10 puan ver" mantığında çalışırken,

öğretmenler anlık karar verme yetisiyle öğrencinin doğru yaklaşım kullanmasına önem vererek küçük hataları görmezden gelebilmekte ve çok boyutlu değerlendirmeler yapabilmektedir. YZ sistemleri "doğru" veya "yanlış" kategorilerinde (1 – 0) değerlendirme yaparken, öğretmenler "kısmen doğru", "iyi yaklaşım", "kavramsal anlama mevcut" gibi küçük ayrıntıları da değerlendirmelerinde kullanabilmektedir. Bu durum, özellikle çok aşamalı problemlerde ve bağlamsal anlama gerektiren sorularda belirgin farklılıklara yol açmaktadır. Soruların puanlanmasında bu tip sorunların aşılması için kullanılan dereceli puanlama anahtarının da YZ sistemleri açısından yetersiz kaldığı görülmektedir. Dereceli puanlama anahtarında önceden belirlenen sınırlı senaryolar ve sabit puanlar YZ sistemleri için bağlayıcı iken, öğretmenlerin anlık ve duruma yönelik karar verme yetisi, öğrencilerin psikolojik olarak yaşadığı kaygıları görebilmesi ve yaratıcı çözümleri fark edebilmesi puanlamalarını daha gerçekçi kılmaktadır. Ayrıca puanlamalardaki farklar ile ilgili sorulan sorularla, YZ sistemlerinin kendi sınırlılıklarını teorik olarak tanımlayabildikleri, ancak pratikte bu farkındalığı uygulamalı değerlendirmelere yansıtamadıkları anlaşılmıştır. Bu durum, YZ sistemlerinin öz farkındalığı olmasına rağmen bunu pratik performansa dönüştüremediklerini göstermektedir ve YZ sistemlerinin mevcut teknik sınırlarını vurgulamaktadır.

Nicel analizlerin yanı sıra yapılan anket, bireysel görüşme ve odak grup görüşmeleri; öğretmenlerin, ölçme uzmanlarının ve öğrencilerin YZ sistemleri ile ilgili görüşlerini ve beklentilerini göstermiştir. Öğretmenlerin yanıtladıkları anket, temkinli ama açık bir yaklaşım sergilediklerini göstermektedir. Ankete katılan öğretmenlerin büyük çoğunluğu YZ sistemlerinin entegrasyonunu desteklemiştir. Görüşme sorularına verilen yanıtlarda ise, öğretmenlerin psikoloji ve empati konularında YZ sistemlerinin sınırlılıklarını vurgulamışlardır.

Öğrenci odak grup görüşmeleri, başarı seviyesine göre farklı bakış açıları olduğunu göstermiştir. Üst grup öğrencilerin büyük çoğunluğu YZ sistemlerinin tarafsız olduğunu belirtirken, alt grup öğrencilerin büyük çoğunluğu bunu anlayışsızlık olarak değerlendirmiştir. Bu durum, YZ sistemlerinin objektifliğinin, farklı başarı seviyelerindeki öğrenciler tarafından farklı algılandığını göstermektedir. Öğretmen ve öğrencilerin kesin olarak anlaştıkları konu ise hibrit bir modelin gerekliliğidir. Üst düzey başarı seviyesindeki öğrenciler "ikisi birbirlerini tamamlıyorlar" görüşünü benimserken, öğretmenlerin çoğunluğu YZ sistemlerinin Türkiye açısından teknik ve pedagojik iyileştirmelerle eğitimde kullanılması gerektiğini belirtmektedir.

Öneriler

YZ ile ilgili uygulama yapacak olan uygulayıcılar; politika yapıcılar ve benzer bir konuda çalışmak isteyen araştırmacılar için öneriler bu başlık altında toplanmıştır.

Uygulayıcılara Yönelik Öneriler

Öğrenciler açısından bakıldığında başarı seviyelerine göre yaklaşımlarının farklı olmasından dolayı, YZ - öğretmen hibrit modeli bu öğrenci gruplarına uygun planlamalar ile eğitim sistemine eklemlenebilir. Üst başarı grubundaki öğrenciler YZ sistemlerini “objektif” olarak algıladıklarından dolayı sistemden alınan verilerin daha fazla kullanıldığı grup olarak belirlenebilirler. Alt başarı grubundaki öğrenciler için ise öğretmen müdahalesi artırılarak YZ sisteminin kontrol edilmesi ve daha fazla insan faktörü olan geri bildirimler verilmesi sağlanabilir. YZ sisteminin değerlendirmeleri her grup öğrenci için gözden geçirilerek, öğrenci ihtiyaçlarına göre esnek puanlama yapılması eğitim sistemine entegrasyonu kolaylaştırabilir.

Teknik ve pedagojik iyileştirme gerekliliği nedeniyle aşamalı entegrasyon planları geliştirilmelidir. Öğretmenlerin bir diğer kaygısı olan altyapı eksiklikleri durumu dikkate alınarak, teknolojik altyapının güçlendirilmesi için belirlenen pilot okullarda entegrasyon çalışmaları yapılarak yaygınlaştırma sürecine geçilmesi önerilir.

Öğretmenlerin ve ölçme uzmanlarının YZ sistemlerini derslerine ve ölçme süreçlerine dahil etmesi; gün geçtikçe gelişen sistemlerle elde edilen hızlı ve kaliteli değerlendirmeleri kullanması kaçınılmaz görünmektedir ancak bu yalnızca gelişen sistemlerin onların kullanımına sunulması ile yeterli bir entegrasyon olmayacaktır. Öğretmenlere yapay zekaya dayalı sistemlerin geliştirilmesi konusunda eğitim verilmesi gereklidir. YZ sistemleri ders içeriklerine ve ölçme süreçlerine eklemlenirken, özellikle karmaşık problemler ve açık uçlu sorular için öğretmen kontrolü sağlanmalı, tüm öğrenci ve öğretmenlerin belirttiği şekilde hibrit model benimsenmelidir. Günlük çalışmalar ve süreç izleme amaçlı yapılan sınıf içi ölçmeler için YZ sistemlerini kullanarak öğretmenlerin sisteme alışması sağlanabilir. Bunun yanında sonuç değerlendirmeye yönelik başarı testlerinde öğretmen kontrolünü devam ettirerek YZ sistemlerinin sınırlılıkları aşılabılır.

Politika Yapıcılara Yönelik Öneriler

Araştırma bulgularına dayalı olarak, MEB YZ kullanımına yönelik ulusal bir strateji geliştirmelidir. Öğretmenlerin YZ destekli eğitim sistemlerini ve teknolojileri kullanabilme yeterliliği, öğretmenlik meslek standartlarına eklenerek eğitimde tüm paydaşların süreç içinde aktif olabilmesi sağlanmalıdır. İçinde bulunan dönemde öğretmenlik yapmakta olan kişiler için hizmetiçi eğitim yapılmasının yanında eğitim fakültelerinde, YZ destekli eğitim sistemlerinin kullanımı ve bu teknolojilerin eğitimini içeren dersler zorunlu hale getirilmelidir. Ayrıca araştırmanın sınırlılıkları dikkate alınarak, uzun süreli, çok disiplinli araştırma projelerine destek verilmelidir. Farklı sosyoekonomik düzeylerde ve farklı sınıf seviyelerinde öğrencilerin bulunduğu pilot okullar belirlenerek YZ'nin öğretmenlerin yerini aldığı değil; derslere ve ölçme süreçlerine destek olduğu stratejiler planlanmalıdır.

Araştırmacılara Yönelik Öneriler

Öğrencilerin ve öğretmenlerin YZ sistemlerine adaptasyon süreçlerini ve YZ sistemlerinin eğitim alanında kullanımının uzun vadeli etkilerini incelemek için en az 2-3 yıllık boylamsal çalışmalar planlanabilir.

Başarı seviyesine göre oluşturulan gruplar “YZ'nin objektif olması” ya da “net sonuçlar vermesi” gibi durumları farklı yorumlamıştır. Bu grupta yerine başarı dışında bir kritere göre öğrencilerin YZ sistemleriyle etkileşimlerini inceleyen araştırmalar yapılabilir.

Araştırmanın nitel analizinde hem öğrencilerden hem de öğretmenlerden alınan veriler doğrultusunda en büyük eksiğin öğrencilerin duygusal durumlarını algılayabilen ve buna göre geri bildirim verebilen sistemlerin geliştirilmesi istenmektedir. Bu alanda yapılan çalışmalar incelenebilir.

KAYNAKLAR

- Akan, S. N. (2024). *Eğitimde Ölçme ve Değerlendirme Aşamasında Yapay Zeka Destekli Uygulamaların İncelenmesi Üzerine İçerik Analizi* [Yüksek lisans tezi, İbn Haldun Üniversitesi]. İbn Haldun Üniversitesi Eğitim Bilimleri Enstitüsü.
- Akdeniz, M., & Özdiç, F. (2021). Eğitimde Yapay Zeka Konusunda Türkiye Adresli Çalışmaların İncelenmesi. *Van Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi*, 18(1), 912-932.
- Alıcı, D.,v.d. (2014). *Eğitimde Ölçme ve Değerlendirme*. Pegem Akademi. Ankara.
- Alpaydın, E. (2020). *Makine Öğrenmesi* (5. baskı). Boğaziçi Üniversitesi Yayınevi. İstanbul.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Arf, C. (1959). *Makine Düşünebilir mi ve Nasıl Düşünebilir? Üniversite Çalışmalarını Muhite Yayma ve Halk Eğitimi Yayınları Konferanslar Serisi*, 1. Erzurum.
- Arslan, K. (2020). Eğitimde Yapay Zeka ve Uygulamaları. *Batı Anadolu Eğitim Bilimleri Dergisi*, 11(1), 71-88.
- Atılğan, H. (2018). *Eğitimde Ölçme ve Değerlendirme*. Anı Yayınları.
- Bahar, M., Nartgün, Z., Durmuş, S., & Bıçak, B. (2008). *Geleneksel Tamamlayıcı Ölçme ve Değerlendirme Teknikleri Öğretmen El Kitabı*. Pegem Akademi. Ankara.
- Bakırcı, Ç. M. (2011). *İnsan Zekasının Evrimi: Neden Sadece İnsanın Beyni Bu Kadar Evrimleşmiştir?*. Evrim Ağacı. <https://evrimagaci.org/s/41>
- Baral, S., Worden, E., Lim, W.-C., Luo, Z., Santorelli, C., Gurung, A., & Heffernan, N. (2025). *Automated feedback in math education: A comparative analysis of LLMs for open-ended responses*. In Proceedings of the 17th International Conference on Educational Data Mining (pp. 80). International Educational Data Mining Society.
- Baykul, Y. (2021). *Eğitimde ve Psikolojide Ölçme: Klâsik Test Teorisi ve Uygulaması*. Pegem Akademi. Ankara.
- Baykul, Y., & Turgut, M. F. (2019). *Eğitimde Ölçme ve Değerlendirme*. Pegem Akademi. Ankara.
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). *Language models are few-shot learners*. Advances in neural information processing systems, 33, 1877-1901.

- Büyüköztürk, Ş., Çokluk, Ö., & Köklü, N. (2017). *Sosyal Bilimler İçin İstatistik*. Pegem Akademi. Ankara.
- Büyüköztürk, Ş., Kılıç Çakmak, E., Erkan Akgün, Ö., Karadeniz, Ş., & Demirel, F. (2021). *Eğitimde Bilimsel Araştırma Yöntemleri (31th ed.)*. Pegem Akademi. Ankara.
- Chen, H., Liu, Z., Gong, J., Li, Y., Wang, L., Liu, W., Xu, L., & Zhang, Z. (2024). *Beyond human subjectivity and error: A novel AI grading system*. arXiv.
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage Publications.
- Crocker, L. M., & Algina, J. (1986). *Introduction to classical and modern test theory*. Wadsworth Publishing Company.
- Çetin, M., & Aktaş, A. (2021). Yapay Zeka ve Eğitimde Gelecek Senaryoları. *OPUS International Journal of Society Researches*, 18 (Eğitim Bilimleri Özel Sayısı), 4225-4268.
- Demirkaya, H., & Sarpel, E. (2018). Demirkaya, H., & Sarpel, E. (2018). Eğitim ve Geliştirme Uygulamalarında Yeni Nesil Bilişim Teknolojilerinden Sanal Gerçeklik, Bulut Bilişim ve Yapay Zeka. *Karadeniz Uluslararası Bilimsel Dergi (40)*, 231-245. <https://doi.org/10.17498/kdeniz.460145>
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). *BERT: Pre-training of deep bidirectional transformers for language understanding*. arXiv preprint arXiv:1810.04805.
- Doğan, C. D., & Aybek, E. C. (2021). *R Shiny ile Psikometri ve İstatistik Uygulamaları*. Pegem Akademi. Ankara. pp.400
- Downing, S. M., & Haladyna, T. M. (2006). *Handbook of test development*. Routledge.
- Dumas, D., Organisciak, P., & Doherty, M. (2021). Measuring divergent thinking originality with human raters and text-mining models: A psychometric comparison of methods. *Psychology of aesthetics, creativity, and the arts*, 15(4), 645–663. <https://doi.org/10.1037/aca0000319>
- ELS Educational Services, Inc. (n.d.). *About ELS*. Retrieved May 8, 2025, from <https://www.els.edu>
- Encyclopædia Britannica. (t.y.). Britannica. 20 Mart 2025 tarihinde <https://www.britannica.com/> adresinden erişildi.
- Erkan, S., Gömleksiz, M., & Karaca, E. (2008). *Eğitimde Ölçme ve Değerlendirme*. Nobel. Ankara.
- Erkuş, A. (2019). *Psikolojide Ölçme ve Ölçek Geliştirme-I Temel Kavramlar ve İşlemler*. Pegem Akademi. Ankara

- Ertürk, S. (2013). *Eğitimde “Program” Geliştirme*. Edge. Ankara.
- Evans, S. W., Williams, A. L., Schultz, B., & Weist, M. D. (2004). *Behavioral assessment in schools*. In C. D. Spielberger (Ed.), *Encyclopedia of Applied Psychology* (pp. 267-275). Elsevier.
- Fielden, J. D. G., & Gibbons, J. D. (1993). *Nonparametric measures of association*. SAGE.
- Gümüş, B. (1976). *Eğitimde Ölçme ve Değerlendirme*. Kâzım Karabekir Eğitim Enstitüsü.
- Gwet, K. L. (2014). *Handbook of inter-rater reliability: the definitive guide to measuring the extent of agreement among raters* (4th ed.). Advanced Analytics, LLC.
- Harari, Y. N. (2015). *Hayvanlardan Tanrılara: Sapiens : İnsan Türünün Kısa Bir Tarihi*. Kolektif Kitap. İstanbul.
- Irmak, H. (2023). *Yapay Zeka Yöntemleri ile Uzaktan Eğitimdeki Sorunların Tespiti ve Öğrencilerin Akademik Performanslarının Tahmin Edilmesi*. İstanbul Üniversitesi.
- İşler, B., & Kılıç, M. (2021). Eğitimde Yapay Zekâ Kullanımı ve Gelişimi. *Yeni Medya Elektronik Dergisi*, 5(1), 1-11.
- Kara, A. (2025). *Açık Uçlu Soruların Yapay Zeka Destekli Otomatik Değerlendirilmesi ve Kullanıcı Deneyimleri* [Doktora tezi, Atatürk Üniversitesi]. Atatürk Üniversitesi Eğitim Bilimleri Enstitüsü.
- Karakaya, İ., Kutlu, Ö., Erman Aslanoğlu, A., & Şata, M. (2022). *Açık Uçlu Soruların Hazırlanması Uygulanması ve Değerlendirilmesi*. Pegem Akademi. Ankara.
- Karasar, N. (2012). *Bilimsel Araştırma Yöntemi: Kavramlar, İlkeler, Teknikler*. Nobel. Ankara.
- Kazu, İ. Y., & Özdemir, O. (2009). Öğrencilerin Bireysel Özelliklerinin Yapay Zeka ile Belirlenmesi (Bulanık Mantık Örneği). *Akademik Bilişim*, 11-13.
- Kelecioğlu, H. & Göçer Şahin, S. (2014). Geçmişten Günümüze Geçerlik. *Journal of Measurement and Evaluation in Education and Psychology*, 5 (2) , 1-11 . DOI: 10.21031/Epod.41706
- Kutlu, Ö., Doğan, C. D., & Karakaya, İ. (2017). *Ölçme ve Değerlendirme Performansa ve Portfolyoya Dayalı Durum Belirleme*. Pegem Akademi. Ankara.
- Landis, J. R., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.
- Landauer, T. K., Laham, D., & Foltz, P. W. (2003). *Automated scoring and annotation of essays with the intelligent essay assessor*. In M. D. Shermis & J. C. Burstein

- (Eds.), *Automated essay scoring: a cross-disciplinary perspective* (pp. 87–112). Lawrence Erlbaum Associates.
- Lu, X., Li, J., Takeuchi, K., & Kashima, H. (2024). AHP-powered LLM reasoning for multi-criteria evaluation of open-ended responses. *Findings of the association for computational linguistics: EMNLP 2024*.
- Martínez-Comesaña, M., Rigueira-Díaz, X., Larrañaga-Janeiro, A., Martínez, J. M., Ocarranza-Prado, I., & Kreibel, D. (2023). *Impact of artificial intelligence on assessment methods in primary and secondary education: Systematic literature review*. *Revista De Psicodidáctica (English Ed.)*, 28(2), 93–103.
- McNichols, H., Zhang, M., & Lan, A. (2023). *Algebra error classification with large language models*. arXiv preprint. <https://doi.org/10.48550/arXiv.2305.06163>
- McNichols, H., Lee, J., Fancsali, S., Ritter, S., & Lan, A. (2024). *Can large language models replicate ITS feedback on open-ended math questions?* arXiv preprint. <https://doi.org/10.48550/arXiv.2405.06414>
- Millî Eğitim Bakanlığı. (2024). *Ortaokul Matematik Dersi Öğretim Programı (5., 6., 7. ve 8. sınıflar)*, Türkiye Yüzyılı Maarif Modeli. T.C. Millî Eğitim Bakanlığı.
- Özçelik, D. A. (2013). *Ölçme ve Değerlendirme Uygulamaları Öğretmen El Kitabı*. Pegem Akademi. Ankara.
- Özçelik, D. A. (2016). *Ölçme ve Değerlendirme*. Pegem Akademi. Ankara.
- Özdemir, S. M. (2010). *İlköğretim Öğretmenlerinin Alternatif Ölçme ve Değerlendirme Araçlarına İlişkin Yeterlilikleri ve Hizmet İçi Eğitim İhtiyaçları*. dergipark.org.
- Özgür, S. B. (2021). *Algoritmalar, yapay zeka, makine öğrenmesi, derin öğrenme ve uygulamaları: Beşeri fayda üretiminin yazılımlar tarafından karşılanması*. *Ekonomi ve Yönetim Araştırmaları Dergisi*, 10(1), 1-29.
- Özgüven, İ. E. (2007). *Psikolojik Testler*. Nobel. Ankara.
- Page, E. B. (1966). *The imminence of grading essays by computer*. *The Phi Delta Kappan*, 47(5), 238–243. <https://www.jstor.org/stable/20342692>
- Popham, W. J. (2000). *Modern educational measurement: Practical guidelines for educational leaders* (3rd ed.).
- Raz, T., Luchini, S., Beaty, R. E., & Kenett, Y. N. (2024). *Automated scoring of open-ended question complexity: A large language model approach*. (Research Square), 1-25. <https://doi.org/10.21203/rs.3.rs-3890828/v1>
- Resmî Gazete. (2023, 9 Eylül). 2023/09/09-2. <https://www.resmigazete.gov.tr/eskiler/2023/09/20230909-2.htm>

- Rogers, A., Kovaleva, O., & Rumshisky, A. (2020). *A primer on neural network models for natural language processing*. Journal of Artificial Intelligence Research, 57, 615-732.
- Say, C. (2018). *50 Soruda Yapay Zeka*. Bilim ve Gelecek Kitaplığı.
- Selçuk, Z. (2009). *Eğitim Psikolojisi*. Nobel. Ankara.
- Skinner, B. (1968). *The technology of teaching*. New York: Appleton-Century-Crofts.
- Sönmez, V. (2011). *Öğretim İlke ve Yöntemleri*. Anı Yayınları. Ankara.
- Stevens, D. D., & Levi, A. (2005). *Introduction to rubrics: An assessment tool to save grading time, convey effective feedback, and promote student learning*. Stylus Pub Llc.
- Tekin, H. (1996). *Eğitimde Ölçme ve Değerlendirme*. Yargı. Ankara.
- Yalçın-Çelik, A. & K.Çoban, Ö. (2023). Kimya Sorularının Cevaplanmasında Yapay Zekâ Tabanlı Sohbet Robotlarının Performansının İncelenmesi. *Türk Eğitim Bilimleri Dergisi*, 21(3), 1540-1561.
- Yılmaz, K. (2025). *Açık Uçlu Maddelerin Otomatik Puanlanmasında Doğal Dil İşleme Yönteminden Yararlanılarak Makine Öğrenmesi Algoritmalarının Karşılaştırılması* [Doktora tezi, Ankara Üniversitesi]. Ankara.
- Wang, R., Wang, R., Shen, Y., Wu, C., Zhou, Q., & Chandra, R. (2025). *Evaluation of LLMs for mathematical problem solving*. arXiv preprint. <https://doi.org/10.48550/arXiv.2506.00309>
- Williamson, D. M., Xi, X., & Breyer, F. J. (2012). *A framework for evaluation and use of automated scoring*. ETS Research Report Series, 2012(1), i-50. <https://doi.org/10.1002/j.2333-8504.2012.tb02387.x>
- Zhao, C., Silva, M., & Poulsen, S. (2025). *Language models are few-shot graders*. University of Illinois Urbana-Champaign and Utah State University.
- <https://claude.ai/>
- <https://odsgm.meb.gov.tr/>
- <https://sozluk.gov.tr/>
- <https://www.resmigazete.gov.tr/eskiler/2023/09/20230909-2.htm>



EK 1. 2023-2024 Eğitim Öğretim Yılı 7. Sınıf 2. Dönem Matematik Yazılısı

Soruları

Soru 1: “Bir sayının 5 katı ile aynı sayının 3 fazlasının farkı 12’ye eşittir.”
Yukarıdaki ifadeye ait denklemini kurunuz.(10 puan)

Soru 2: Aşağıdaki denklemlerde bilinmeyen alacağı değerleri bulunuz. (2x5 puan = 10 puan)

a) $3x - 7 =$ b) $7x + 2(2 - x) =$

Soru 3: Aşağıda 2085 m uzunluğunda düz bir yol üzerine eşit aralıklarla yerleştirilen reklam panoları gösterilmiştir.



Yolun kenarına 50 m aralıklarla 37 adet reklam panosu yerleştirilmiştir.
Yolun başında ve sonunda 50 metre boşluk olduğuna göre yerleştirilen panolardan bir tanesinin genişliği (x) kaç metredir? (10 puan)

Soru 4: Aşağıda verilen orantılarda x değerlerini bulunuz.(2x5 puan = 10 Puan)

a) $4/7 = 12/$ b) $3/7 = (x-1)/$

Soru 5: Aşağıda verilen tabloda kolilerde bulunan paket sayıları verilmiştir. (2x5 = 10 puan)

Tablo: Kolilerde Bulunan Paket Sayısı

Koli Sayısı	3	1	2	z	6
Paket sayısı	15	x	y	20	30

Yukarıdaki tabloya göre

a) Koli sayısı ile paket sayısı arasındaki orantı çeşidini yazarak, orantı sabitini bulunuz.

b) $x + y + z$ işleminin sonucunu bulunuz.

Soru 6: Bir makinenin birbirini döndüren iki dişli çarkından küçük olanında 16 dişli, büyük olanında 24 dişli vardır. Buna göre küçük dişli 3 tam tur döndüğünde büyük dişli kaç tam tur dönüş yapar?

(10 puan)

Soru 7 : Bir yerin, bir arazinin veya bölgenin belli bir oranda küçültülerek kâğıt üzerine çizilmiş şekline plan denir. Bir planda gerçek uzunlukları küçültme oranına ölçek denir.

İki ev arası, ölçeği $1/14000$ olan mahalle planında 5 cm ile gösterilmiştir. Bu iki ev arası mesafe gerçekte kaç metredir?

(1 m = 100 cm) (10 puan)

Soru 8 : Emre bir kitabın önce %10'unu, daha sonra kalanın %20'sini okuyor.
Geriye 288 sayfa kaldığına göre Emre'nin okuduğu kitap kaç sayfadır? (10 puan)

Soru 9 : Aşağıdaki soruları yanıtlayınız. (2x5 puan =10 Puan)

- a) 80 sayısının % 20'si kaçtır? b) % 5'i 12 olan sayının tamamı kaçtır?

Soru 10 : Aşağıdaki soruları yanıtlayınız.(2x5 puan = 10 Puan)

- a) % 40 eksik 60 olan sayı kaçtır b) 30'un % 25 fazlası kaçtır?



EK 2. Dereceli Puanlama Anahtarı

1. soru

Tam Doğru Yanıt (TD)		
TD1	$5x - (x + 3) = 12$ [x veya başka bir bilinmeyen sembolü]	10 puan
TD2	$5x - x - 3 = 12$ [x veya başka bir bilinmeyen sembolü]	10 puan
Kısmen Doğru Yanıtlar (KD)		
KD1	$5x - (x + 3)$ [eşitliği tamamlamamış]	9 puan
KD2	$5x - x + 3 = 12$ [parantez kullanmamış]	8 puan
KD3	$5x + x + 3 = 12$ [işlem hatası]	5 puan
KD4	$5x - 3 = 12 / 5x + 3 = 12$ [soruyu eksik anlamış]	2 puan
Yanlış Yanıtlar (YY)		
YY1	TD ve KD dışındaki yanıtlar	0 puan
Boş Yanıt (BY)		
BY1	İşlem yapılmamış – boş bırakılmış	0 puan

2. soru a kısmı

Tam Doğru Yanıt (TD)		
TD1	$3x - 7 = 17$ $3x = 17 + 7$ [ya da $3x = 24$] $x = 24/3$ $x = 8$	5 puan
Kısmen Doğru Yanıtlar (KD)		
KD1	$x = 24/3$ $x = 8$ [tam işlemi göstermemiş]	4 puan
KD2	$x = 8$ [tam işlemi göstermemiş]	4 puan
Yanlış Yanıtlar (YY)		
YY1	TD ve KD dışındaki yanıtlar	0 puan
Boş Yanıt (BY)		
BY1	İşlem yapılmamış – boş bırakılmış	0 puan

2. soru b kısmı

Tam Doğru Yanıt (TD)		
----------------------	--	--

TD1	$7x + 2(2 - x) = 19$ $7x + 4 - 2x = 19$ $5x + 4 = 19$ $5x = 19 - 4$ [ya da $5x = 15$] $x = 3$	5 puan
Kısmen Doğru Yanıtlar (KD)		
KD1	$7x + 4 - 2x = 19$ $x = 3$ [tam işlemi göstermemiş]	4 puan
KD2	$x = 3$ [tam işlemi göstermemiş]	4 puan
Yanlış Yanıtlar (YY)		
YY1	TD ve KD dışındaki yanıtlar	0 puan
Boş Yanıt (BY)		
BY1	İşlem yapılmamış – boş bırakılmış	0 puan

3. soru

Tam Doğru Yanıt (TD)		
TD1	$38 \cdot 50 + 37 \cdot x = 2085$ $37x = 2085 - 1900$ $37x = 185$ $x = 5$	Bu işlemleri içeren her yanıt 10 puan
Kısmen Doğru Yanıtlar (KD)		
KD1	$50 \cdot 38 = 1900$ $2085 - 1900 = 185$ $185 / 37 = 5$ [Denklemini kurmamış]	9 puan
KD2	$x = 5$ [işlem yapmamış]	5 puan
KD3	$37 \cdot 50 = 1850$ $2085 - 1850 = 235$ [aralıkları yanlış hesaplamış]	3 puan
Yanlış Yanıtlar (YY)		
YY1	TD ve KD dışındaki yanıtlar	0 puan
Boş Yanıt (BY)		
BY1	İşlem yapılmamış – boş bırakılmış	0 puan

4. soru a kısmı

Tam Doğru Yanıt (TD)		
TD1	$\frac{4}{7} = \frac{12}{x}$ $4x = 12 \cdot 7 = 84$	5 puan

	$x = 84 / 4$ $x = 21$	
TD2	$x / 7 = 12 / 4$ $x / 7 = 3$ $x = 3 \cdot 7 = 21$	5 puan
TD3	$\frac{4(\cdot 3)}{7(\cdot 3)} = \frac{12}{x}$ $4(\cdot 3) = 12$ ise $7(\cdot 3) = x$ $x = 21$	5 puan
Kısmen Doğru Yanıtlar (KD)		
KD1	$x = 21$ [tam işlemi göstermemiş]	4 puan
KD2	[Tüm işlemleri yapıp sonucu yanlış bulmuş]	3 puan
Yanlış Yanıtlar (YY)		
YY1	TD ve KD dışındaki yanıtlar	0 puan
Boş Yanıt (BY)		
BY1	İşlem yapılmamış – boş bırakılmış	0 puan

4. soru b kısmı

Tam Doğru Yanıt (TD)		
TD1	$\frac{3}{7} = \frac{x-1}{14}$ $7 \cdot (x - 1) = 3 \cdot 14$ $7x - 7 = 42$ $7x = 49$ $x = 7$	5 puan
TD2	$\frac{3(\cdot 2)}{7(\cdot 2)} = \frac{x-1}{14}$ $7(\cdot 2) = 14$ ise $3(\cdot 2) = x - 1$ $x = 7$	5 puan
Kısmen Doğru Yanıtlar (KD)		
KD1	$7 \cdot (x - 1) = 3 \cdot 14$ $7x - 7 = 42$ $7x = 49$ $x = 7$	4 puan
KD2	$x = 7$ [tam işlemi göstermemiş]	4 puan

KD3	[Tüm işlemleri yapıp sonucu yanlış bulmuş]	3 puan
KD4	$7 \cdot (x - 1) = 3 \cdot 14$ $7x - 1 = 42$ $7x = 43$ $x = 43/7$ [Parantezi açarken işlem hatası]	2 puan
Yanlış Yanıtlar (YY)		
YY1	TD ve KD dışındaki yanıtlar	0 puan
Boş Yanıt (BY)		
BY1	İşlem yapılmamış – boş bırakılmış	0 puan

5. soru a kısmı

Tam Doğru Yanıt (TD)		
TD1	Doğru orantı (koli sayısı / paket sayısı) $3 / 15 = 1 / 5$ [ya da $6 / 30 = 1 / 5$] sabit	5 puan
TD2	Doğru orantı (paket sayısı / koli sayısı) $15 / 3 = 5$ [ya da $30 / 6 = 5$] sabit	5 puan
Yanlış Yanıtlar (YY)		
YY1	TD dışındaki yanıtlar	0 puan
Boş Yanıt (BY)		
BY1	İşlem yapılmamış – boş bırakılmış	0 puan

5. soru b kısmı

Tam Doğru Yanıt (TD)		
TD1	$1 / x = 1 / 5$ $x = 5$ $2 / y = 1 / 5$ $y = 10$ $z / 20 = 1 / 5$ $z = 4$ $x + y + z = 19$	5 puan
Kısmen Doğru Yanıtlar (KD)		
KD1	[Tüm işlemleri yapıp sonucu yanlış bulmuş]	4 puan

KD2	[x, y ya da z'den birisi eksik]	2 puan
Yanlış Yanıtlar (YY)		
YY1	TD ve KD dışındaki yanıtlar	0 puan
Boş Yanıt (BY)		
BY1	İşlem yapılmamış – boş bırakılmış	0 puan

6. soru

Tam Doğru Yanıt (TD)		
TD1	$24 \cdot x = 16 \cdot 3$ $24x = 48$ $x = 2$	10 puan
Kısmen Doğru Yanıtlar (KD)		
KD1	[Tüm işlemleri yapıp sonucu yanlış bulmuş]	8 puan
KD2	$48 / 24 = 2$ [Denklem kurmamış, işlem yapmamış]	4 puan
Yanlış Yanıtlar (YY)		
YY1	TD ve KD dışındaki yanıtlar	0 puan
Boş Yanıt (BY)		
BY1	İşlem yapılmamış – boş bırakılmış	0 puan

7. soru

Tam Doğru Yanıt (TD)		
TD1	Krokide ↔ Gerçekte 1 cm ↔ 14000 cm 5 cm ↔ ? cm $14000 \times 5 = 70000$ cm 70000 cm = 700 m	10 puan
Kısmen Doğru Yanıtlar (KD)		
KD1	Krokide ↔ Gerçekte 1 cm ↔ 14000 cm 5 cm ↔ ? cm $14000 \times 5 = 70000$ cm [metre'ye çevirmemiş]	6 puan
KD2	[Tüm işlemler var ancak işlem hatası yapmış]	4 puan
Yanlış Yanıtlar (YY)		

YY1	TD ve KD dışındaki yanıtlar	0 puan
Boş Yanıt (BY)		
BY1	İşlem yapılmamış – boş bırakılmış	0 puan

8. soru

Tam Doğru Yanıt (TD)		
TD1	$\frac{\text{Önce}}{\frac{10}{100}} + \left(\frac{\text{Sonra}}{\frac{100}{100} - \frac{10}{100}} \right) \cdot \frac{20}{100} = \frac{\text{Toplam}}{\frac{28}{100}} \quad \frac{\text{Kalan}}{\frac{100}{100} - \frac{28}{100}} = \frac{72}{100}$ Oran ↔ Kitap %72 ↔ 288 sayfa %100 ↔ ? sayfa $100 \cdot 288 = 28800$ $28800 / 72 = 400$ sayfa	10 puan
Kısmen Doğru Yanıtlar (KD)		
KD1	[İşlem basamakları doğru sonuç hatalı]	8 puan
KD2	[Orantı yanlış kurulmuş, işlemler doğru]	4 puan
Yanlış Yanıtlar (YY)		
YY1	TD ve KD dışındaki yanıtlar	0 puan
Boş Yanıt (BY)		
BY1	İşlem yapılmamış – boş bırakılmış	0 puan

9. soru a kısmı

Tam Doğru Yanıt (TD)		
TD1	Oran ↔ Sayı %100 ↔ 80 %20 ↔ ? $(20 \cdot 80) / 100 = 16$	5 puan
Kısmen Doğru Yanıtlar (KD)		
KD1	[Tüm işlemleri yapıp sonucu yanlış bulmuş]	3 puan
KD2	$(20 \cdot 80) / 100 = 16$ [Denklem kurmamış, işlem yapmış]	3 puan
Yanlış Yanıtlar (YY)		

YY1	TD ve KD dışındaki yanıtlar	0 puan
Boş Yanıt (BY)		
BY1	İşlem yapılmamış – boş bırakılmış	0 puan

9. soru b kısmı

Tam Doğru Yanıt (TD)		
TD1	Oran ↔ Sayı %5 ↔ 12 %100 ↔ ? $(12 \cdot 100) / 5 = 240$	5 puan
Kısmen Doğru Yanıtlar (KD)		
KD1	[Tüm işlemleri yapıp sonucu yanlış bulmuş]	3 puan
KD2	$(12 \cdot 100) / 5 = 240$ [Denklem kurmamış, işlem yapmış]	3 puan
Yanlış Yanıtlar (YY)		
YY1	TD ve KD dışındaki yanıtlar	0 puan
Boş Yanıt (BY)		
BY1	İşlem yapılmamış – boş bırakılmış	0 puan

10. soru a kısmı

Tam Doğru Yanıt (TD)		
TD1	$\%100 - \%40 = \%60$ Oran ↔ Sayı %60 ↔ 60 %100 ↔ ? $(100 \cdot 60) / 60 = 100$	5 puan
Kısmen Doğru Yanıtlar (KD)		
KD1	[Tüm işlemleri yapıp sonucu yanlış bulmuş]	3 puan
KD2	$(100 \cdot 60) / 60 = 100$ [Denklem kurmamış, işlem yapmış]	3 puan
Yanlış Yanıtlar (YY)		
YY1	TD ve KD dışındaki yanıtlar	0 puan
Boş Yanıt (BY)		
BY1	İşlem yapılmamış – boş bırakılmış	0 puan

10. soru b kısmı

Tam Doğru Yanıt (TD)		
TD1	$\%100 + \%25 = \%125$ Oran $\leftrightarrow$ Sayı $\%100 \leftrightarrow 30$ $\%125 \leftrightarrow ?$ $(125 \cdot 30) / 100 = 37,5$	5 puan
Kısmen Doğru Yanıtlar (KD)		
KD1	[Tüm işlemleri yapıp sonucu yanlış bulmuş]	3 puan
KD2	$(125 \cdot 30) / 100 = 37,5$ [Denklem kurmamış, işlem yapmış]	3 puan
Yanlış Yanıtlar (YY)		
YY1	TD ve KD dışındaki yanıtlar	0 puan
Boş Yanıt (BY)		
BY1	İşlem yapılmamış – boş bırakılmış	0 puan

EK 3. Başarı Seviyesine Göre Belirlenen Odak Grupların Öğretmen Puanları**Tablosu**

Öğretmen Puanları											
	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	Puan
Oğrenci-Ust-1	10	10	10	10	10	10	10	10	10	10	100
Oğrenci-Ust-2	10	10	10	10	10	10	10	10	10	10	100
Oğrenci-Ust-3	10	10	10	10	10	10	10	10	10	10	100
Oğrenci-Ust-4	10	10	10	10	10	10	10	10	10	10	100
Oğrenci-Ust-5	10	10	10	10	10	10	10	10	10	10	100
Oğrenci-Ust-6	10	10	10	10	10	10	10	10	10	10	100
Oğrenci-Ust-7	10	10	10	10	10	10	10	10	10	10	100
Oğrenci-Ust-8	10	10	10	10	10	10	10	10	10	10	100
Oğrenci-Ust-9	10	10	10	9	9	10	10	10	10	10	98
Oğrenci-Ust-10	10	10	10	10	10	10	7	10	10	10	97
Oğrenci-Orta-1	8	10	10	10	10	10	8	10	10	10	96
Oğrenci-Orta-2	7	10	10	10	10	10	8	10	10	10	95
Oğrenci-Orta-3	7	10	10	10	10	10	8	10	10	10	95
Oğrenci-Orta-4	6	10	10	8	10	10	8	10	10	10	92
Oğrenci-Orta-5	10	10	2	10	10	10	8	10	10	10	90
Oğrenci-Orta-6	8	10	10	8	10	10	10	2	10	10	88
Oğrenci-Orta-7	10	5	10	10	10	2	10	10	10	10	87
Oğrenci-Orta-8	10	8	10	10	10	5	10	8	10	4	85
Oğrenci-Orta-9	8	8	6	9	9	9	10	2	9	10	80
Oğrenci-Orta-10	7	10	10	10	0	10	10	3	10	10	80
Oğrenci-Alt-1	8	5	10	10	10	10	0	0	10	10	73
Oğrenci-Alt-2	10	5	10	9	10	10	6	0	5	5	70
Oğrenci-Alt-3	4	7	5	8	5	10	3	10	7	10	69
Oğrenci-Alt-4	10	8	2	10	6	9	8	3	5	4	65
Oğrenci-Alt-5	6	10	5	8	0	0	5	5	10	7	56
Oğrenci-Alt-6	3	2	10	9	9	3	2	0	10	4	52
Oğrenci-Alt-7	5	10	0	10	2	1	10	10	1	1	50
Oğrenci-Alt-8	6	0	6	9	0	8	0	5	5	7	46
Oğrenci-Alt-9	4	0	4	10	9	0	3	0	3	5	38
Oğrenci-Alt-10	6	0	3	3	0	0	8	2	0	0	22

EK 4. Başarı Seviyesine Göre Belirlenen Odak Grupların YZ Puanları Tablosu

Yapay Zeka Puanları											
	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	Puan
Ogrenci-Ust-1	10	10	10	10	10	10	10	10	10	10	100
Ogrenci-Ust-2	10	10	10	10	10	10	10	10	10	10	100
Ogrenci-Ust-3	10	10	10	10	10	10	10	10	10	10	100
Ogrenci-Ust-4	10	10	10	10	10	10	10	10	10	10	100
Ogrenci-Ust-5	10	10	10	10	10	10	10	10	10	10	100
Ogrenci-Ust-6	10	10	10	10	10	10	10	10	10	10	100
Ogrenci-Ust-7	10	10	10	10	10	10	10	10	10	10	100
Ogrenci-Ust-8	10	10	10	10	10	10	10	10	10	10	100
Ogrenci-Ust-9	10	10	10	10	10	10	10	10	10	10	100
Ogrenci-Ust-10	10	10	10	10	10	10	10	10	10	10	100
Ogrenci-Orta-1	8	8	10	8	10	10	10	10	10	3	87
Ogrenci-Orta-2	4	10	10	10	10	10	4	10	10	10	88
Ogrenci-Orta-3	9	9	9	8	10	10	4	8	10	10	87
Ogrenci-Orta-4	0	10	10	9	10	10	10	10	10	10	89
Ogrenci-Orta-5	10	8	2	10	10	10	5	10	10	10	85
Ogrenci-Orta-6	8	10	10	5	10	10	10	2	10	10	85
Ogrenci-Orta-7	10	8	10	10	10	10	10	2	10	10	90
Ogrenci-Orta-8	10	9	10	10	10	8	10	3	10	10	90
Ogrenci-Orta-9	8	10	6	10	10	10	10	2	10	10	86
Ogrenci-Orta-10	8	10	10	10	5	10	10	2	10	10	85
Ogrenci-Alt-1	0	6	9	7	7	4	4	3	6	6	52
Ogrenci-Alt-2	8	10	10	10	10	10	3	0	4	5	70
Ogrenci-Alt-3	2	7	3	8	7	10	4	2	10	10	63
Ogrenci-Alt-4	8	6	2	6	6	7	4	3	5	4	51
Ogrenci-Alt-5	0	10	0	4	0	8	0	0	10	5	37
Ogrenci-Alt-6	2	3	1	7	6	2	0	1	10	9	41
Ogrenci-Alt-7	0	2	0	3	2	1	2	10	1	1	22
Ogrenci-Alt-8	2	2	5	6	0	2	0	7	4	7	35
Ogrenci-Alt-9	2	7	3	10	8	5	0	2	7	5	49
Ogrenci-Alt-10	2	0	3	0	0	0	0	0	0	3	8

EK 5. Etik Kurul Raporu

ANKARA ÜNİVERSİTESİ ALT ETİK KURULU KARAR ÖRNEĞİ

Karar Tarihi : 06.05.2024

Toplantı Sayısı : 14

Karar Sayısı : 164

164- Üniversitemiz Eğitim Bilimleri Fakültesi Eğitim Bilimleri Bölümü öğretim üyesi Doç. Dr. Seher YALÇIN'ın danışmanlığını yaptığı, yüksek lisans öğrencisi Serra Sezin ŞENER'in "Ortaokul 7. Sınıf Matematik Dersinde Açık Uçlu Soruların Değerlendirilmesi Yapay Zekâ ve Öğretmen Puanlamalarının Karşılaştırılması" başlıklı tezi ile ilgili 06.05.2024 tarihli "İnsan Üzerinde Yapılan Klinik Dışı Araştırmalar Başvuru Formu" Etik Kurulumuzca incelendi.

Üniversitemiz Eğitim Bilimleri Fakültesi Eğitim Bilimleri Bölümü öğretim üyesi Doç. Dr. Seher YALÇIN'ın danışmanlığını yaptığı, yüksek lisans öğrencisi Serra Sezin ŞENER'in " Ortaokul 7. Sınıf Matematik Dersinde Açık Uçlu Soruların Değerlendirilmesi Yapay Zekâ ve Öğretmen Puanlamalarının Karşılaştırılması " başlıklı tezi ile ilgili araştırma protokolüne uyulması ve etik onay tarihinden itibaren geçerli olması koşuluyla uygulanmasının etik açıdan uygun olduğuna oy birliği ile karar verilmiştir.

**ASLININ AYNIDIR
06/05/2024**

Prof. Dr. Muharrem ÖZEN
Ankara Üniversitesi
Etik Kurulu Başkanı

EK 6. Okul Müdürleri Onam Formları

Aydınlatılmış Onam Formu

Sayın Okul Müdürü,

“Ortaokul 7. Sınıf Matematik Dersinde Açık Uçlu Soruların Değerlendirilmesi Yapay Zekâ ve Öğretmen Puanlamalarının Karşılaştırılması” adlı araştırmaya katılmak üzere davet edilmiş bulunuyorsunuz. Bu araştırmada yer almayı kabul etmeden önce, araştırmanın ne amaçla yapılmak istendiğini anlamanız ve kararınızı bu bilgilendirme çerçevesinde özgürce vermeniz gerekmektedir. Aşağıdaki bilgileri lütfen dikkatlice okuyunuz, sorularınız olursa sorunuz ve açık yanıtlar isteyiniz.

Bu araştırmanın amacı sınıf içi ölçmelerde kullanılan açık uçlu maddelerden oluşan bir matematik testinde, Yapay Zeka (YZ) ve öğretmen değerlendirmelerinin etkinliğini ve güvenilirliğini karşılaştırmak; yöntemlerin test istatistikleri ve sonuç raporları arasındaki tutarlılıkları ve farklılıklarını incelemek ve her öğrenciye özelleştirilmiş geri bildirim sağlayan karnelerin hazırlanmasında YZ ve öğretmen değerlendirmesi yöntemlerinin öğrenci başarısı ve motivasyonu üzerindeki etkilerini analiz etmektir. Araştırma için 200 öğrenciden veri toplanması hedeflenmektedir. Araştırma gizlilik esasına dayalı olarak yürütülecektir. Katılımcılardan elde edilen veriler araştırma kapsamı dışında kullanılmayacaktır.

Bu araştırmada yer almak tümüyle sizin isteğinize bağlıdır. Araştırmada yer almayı reddedebilirsiniz ya da başladıktan sonra yarıda bırakabilirsiniz. Bu araştırmanın sonuçları bilimsel amaçlarla kullanılacaktır. Araştırmadan çekilmeniz ya da araştırmacı tarafından araştırmadan çıkarılmanız halinde, sizle ilgili veriler kullanılmayacaktır. Ancak veriler bir kez anonimleştikten sonra araştırmadan çekilmeniz mümkün olmayacaktır. Sizden elde edilen tüm bilgiler gizli tutulacak araştırma yayınlandığında da varsa kimlik bilgilerinizin gizliliği korunacaktır. Ayrıca araştırmada sizden herhangi bir ücret talep edilmeyip, herhangi bir ücret de ödenmeyecektir.

Yukarıda yer alan ve araştırmaya başlanmadan önce gönüllülere verilmesi gereken bilgileri içeren metni okudum (ya da sözlü olarak dinledim). Eksik kaldığını düşündüğüm konularda sorularımı araştırmacılara sordum ve doyurucu yanıtlar aldım. Yazılı ve sözlü olarak tarafıma sunulan tüm açıklamaları ayrıntılarıyla anladığım kanısındayım. Çalışmaya katılmayı isteyip istemediğim konusunda karar vermem için yeterince zaman tanındı. Bu koşullar altında, araştırma kapsamında elde edilen şahsıma ait bilgilerin bilimsel amaçlarla kullanılmasını, gizlilik kurallarına uyulmak kaydıyla sunulmasını ve yayınlanmasını, hiçbir baskı ve zorlama altında kalmaksızın, kendi özgür irademle kabul ettiğimi beyan ederim.

Gönüllünün adı/soyadı

Anıl ŞERAF

İmza/Tarih

20.12.2024/

Araştırmacının adı/soyadı

Serra Sezin Şener

İmza/Tarih

20.12.2024

Aydınlatılmış Onam Formu

Sayın Okul Müdürü,

“Ortaokul 7. Sınıf Matematik Dersinde Açık Uçlu Soruların Değerlendirilmesi Yapay Zekâ ve Öğretmen Puanlamalarının Karşılaştırılması” adlı araştırmaya katılmak üzere davet edilmiş bulunuyorsunuz. Bu araştırmada yer almayı kabul etmeden önce, araştırmanın ne amaçla yapılmak istendiğini anlamanız ve kararınızı bu bilgilendirme çerçevesinde özgürce vermeniz gerekmektedir. Aşağıdaki bilgileri lütfen dikkatlice okuyunuz, sorularınız olursa sorunuz ve açık yanıtlar isteyiniz.

Bu araştırmanın amacı sınıf içi ölçmelerde kullanılan açık uçlu maddelerden oluşan bir matematik testinde, Yapay Zeka (YZ) ve öğretmen değerlendirmelerinin etkinliğini ve güvenilirliğini karşılaştırmak; yöntemlerin test istatistikleri ve sonuç raporları arasındaki tutarlılıkları ve farklılıklarını incelemek ve her öğrenciye özelleştirilmiş geri bildirim sağlayan karnelerin hazırlanmasında YZ ve öğretmen değerlendirmesi yöntemlerinin öğrenci başarısı ve motivasyonu üzerindeki etkilerini analiz etmektir. Araştırma için 200 öğrenciden veri toplanması hedeflenmektedir. Araştırma gizlilik esasına dayalı olarak yürütülecektir. Katılımcılardan elde edilen veriler araştırma kapsamı dışında kullanılmayacaktır.

Bu araştırmada yer almak tümüyle sizin isteğinize bağlıdır. Araştırmada yer almayı reddedebilirsiniz ya da başladıktan sonra yarıda bırakabilirsiniz. Bu araştırmanın sonuçları bilimsel amaçlarla kullanılacaktır. Araştırmadan çekilmeniz ya da araştırmacı tarafından araştırmadan çıkarılmanız halinde, sizle ilgili veriler kullanılmayacaktır. Ancak veriler bir kez anonimleştikten sonra araştırmadan çekilmeniz mümkün olmayacaktır. Sizden elde edilen tüm bilgiler gizli tutulacak araştırma yayınlandığında da varsa kimlik bilgilerinizin gizliliği korunacaktır. Ayrıca araştırmada sizden herhangi bir ücret talep edilmeyip, herhangi bir ücret de ödenmeyecektir.

Yukarıda yer alan ve araştırmaya başlanmadan önce gönüllülere verilmesi gereken bilgileri içeren metni okudum (ya da sözlü olarak dinledim). Eksik kaldığını düşündüğüm konularda sorularımı araştırmacılara sordum ve doyurucu yanıtlar aldım. Yazılı ve sözlü olarak tarafıma sunulan tüm açıklamaları ayrıntılarıyla anladığım kanısındayım. Çalışmaya katılmayı isteyip istemediğim konusunda karar vermeme için yeterince zaman tanındı. Bu koşullar altında, araştırma kapsamında elde edilen şahsıma ait bilgilerin bilimsel amaçlarla kullanılmasını, gizlilik kurallarına uyulmak kaydıyla sunulmasını ve yayınlanmasını, hiçbir baskı ve zorlama altında kalmaksızın, kendi özgür irademle kabul ettiğimi beyan ederim.

Gönüllünün adı/soyadı Fatma KURAT

İmza/Tarih

Araştırmacının adı/soyadı

İmza/Tarih

20.12.2024
Serra Setta Seio
20.12.2024

Aydınlatılmış Onam Formu

Sayın Okul Müdürü,

“Ortaokul 7. Sınıf Matematik Dersinde Açık Uçlu Soruların Değerlendirilmesi Yapay Zekâ ve Öğretmen Puanlamalarının Karşılaştırılması” adlı araştırmaya katılmak üzere davet edilmiş bulunuyorsunuz. Bu araştırmada yer almayı kabul etmeden önce, araştırmanın ne amaçla yapılmak istendiğini anlamanız ve kararınızı bu bilgilendirme çerçevesinde özgürce vermeniz gerekmektedir. Aşağıdaki bilgileri lütfen dikkatlice okuyunuz, sorularınız olursa sorunuz ve açık yanıtlar isteyiniz.

Bu araştırmanın amacı sınıf içi ölçmelerde kullanılan açık uçlu maddelerden oluşan bir matematik testinde, Yapay Zeka (YZ) ve öğretmen değerlendirmelerinin etkinliğini ve güvenilirliğini karşılaştırmak; yöntemlerin test istatistikleri ve sonuç raporları arasındaki tutarlılıkları ve farklılıklarını incelemek ve her öğrenciye özelleştirilmiş geri bildirim sağlayan karnelerin hazırlanmasında YZ ve öğretmen değerlendirmesi yöntemlerinin öğrenci başarısı ve motivasyonu üzerindeki etkilerini analiz etmektir. Araştırma için 200 öğrenciden veri toplanması hedeflenmektedir. Araştırma gizlilik esasına dayalı olarak yürütülecektir. Katılımcılardan elde edilen veriler araştırma kapsamı dışında kullanılmayacaktır.

Bu araştırmada yer almak tümüyle sizin isteğinize bağlıdır. Araştırmada yer almayı reddedebilirsiniz ya da başladıktan sonra yarıda bırakabilirsiniz. Bu araştırmanın sonuçları bilimsel amaçlarla kullanılacaktır. Araştırmadan çekilmeniz ya da araştırmacı tarafından araştırmadan çıkarılmanız halinde, sizle ilgili veriler kullanılmayacaktır. Ancak veriler bir kez anonimleştikten sonra araştırmadan çekilmeniz mümkün olmayacaktır. Sizden elde edilen tüm bilgiler gizli tutulacak araştırma yayınlandığında da varsa kimlik bilgilerinizin gizliliği korunacaktır. Ayrıca araştırmada sizden herhangi bir ücret talep edilmeyip, herhangi bir ücret de ödenmeyecektir.

Yukarıda yer alan ve araştırmaya başlanmadan önce gönüllülere verilmesi gereken bilgileri içeren metni okudum (ya da sözlü olarak dinledim). Eksik kaldığını düşündüğüm konularda sorularımı araştırmacılara sordum ve doyurucu yanıtlar aldım. Yazılı ve sözlü olarak tarafıma sunulan tüm açıklamaları ayrıntılarıyla anladığımı kanıslındayım. Çalışmaya katılmayı isteyip istemediğim konusunda karar vermem için yeterince zaman tanındı. Bu koşullar altında, araştırma kapsamında elde edilen şahsıma ait bilgilerin bilimsel amaçlarla kullanılmasını, gizlilik kurallarına uyulmak kaydıyla sunulmasını ve yayınlanmasını, hiçbir baskı ve zorlama altında kalmaksızın, kendi özgür irademle kabul ettiğimi beyan ederim.

Gönüllünün adı/soyadı

Dilek DAĞYAR

İmza/Tarih

20.12.2024

Araştırmacının adı/soyadı

Serra Serra Seir

İmza/Tarih

20.12.2024

Aydınlatılmış Onam Formu

Sayın Okul Müdürü,

“Ortaokul 7. Sınıf Matematik Dersinde Açık Uçlu Soruların Değerlendirilmesi Yapay Zekâ ve Öğretmen Puanlamalarının Karşılaştırılması” adlı araştırmaya katılmak üzere davet edilmiş bulunuyorsunuz. Bu araştırmada yer almayı kabul etmeden önce, araştırmanın ne amaçla yapılmak istendiğini anlamanız ve kararınızı bu bilgilendirme çerçevesinde özgürce vermeniz gerekmektedir. Aşağıdaki bilgileri lütfen dikkatlice okuyunuz, sorularınız olursa sorunuz ve açık yanıtlar isteyiniz.

Bu araştırmanın amacı sınıf içi ölçmelerde kullanılan açık uçlu maddelerden oluşan bir matematik testinde, Yapay Zeka (YZ) ve öğretmen değerlendirmelerinin etkinliğini ve güvenilirliğini karşılaştırmak; yöntemlerin test istatistikleri ve sonuç raporları arasındaki tutarlılıkları ve farklılıklarını incelemek ve her öğrenciye özelleştirilmiş geri bildirim sağlayan karnelerin hazırlanmasında YZ ve öğretmen değerlendirmesi yöntemlerinin öğrenci başarısı ve motivasyonu üzerindeki etkilerini analiz etmektir. Araştırma için 200 öğrenciden veri toplanması hedeflenmektedir. Araştırma gizlilik esasına dayalı olarak yürütülecektir. Katılımcılardan elde edilen veriler araştırma kapsamı dışında kullanılmayacaktır.

Bu araştırmada yer almak tümüyle sizin isteğinize bağlıdır. Araştırmada yer almayı reddedebilirsiniz ya da başladıktan sonra yarıda bırakabilirsiniz. Bu araştırmanın sonuçları bilimsel amaçlarla kullanılacaktır. Araştırmadan çekilmeniz ya da araştırmacı tarafından araştırmadan çıkarılmanız halinde, sizle ilgili veriler kullanılmayacaktır. Ancak veriler bir kez anonimleştikten sonra araştırmadan çekilmeniz mümkün olmayacaktır. Sizden elde edilen tüm bilgiler gizli tutulacak araştırma yayınlandığında da varsa kimlik bilgilerinizin gizliliği korunacaktır. Ayrıca araştırmada sizden herhangi bir ücret talep edilmeyip, herhangi bir ücret de ödenmeyecektir.

Yukarıda yer alan ve araştırmaya başlanmadan önce gönüllülere verilmesi gereken bilgileri içeren metni okudum (ya da sözlü olarak dinledim). Eksik kaldığını düşündüğüm konularda sorularımı araştırmacılara sordum ve doyurucu yanıtlar aldım. Yazılı ve sözlü olarak tarafıma sunulan tüm açıklamaları ayrıntılarıyla anladığım kanısındayım. Çalışmaya katılmayı isteyip istemediğim konusunda karar vermem için yeterince zaman tanındı. Bu koşullar altında, araştırma kapsamında elde edilen şahsıma ait bilgilerin bilimsel amaçlarla kullanılmasını, gizlilik kurallarına uyulmak kaydıyla sunulmasını ve yayınlanmasını, hiçbir baskı ve zorlama altında kalmaksızın, kendi özgür irademle kabul ettiğimi beyan ederim.

Gönüllünün adı/soyadı

Özlem KONAK
/ 31.12.2024

İmza/Tarih

Araştırmacının adı/soyadı

Serra Setra Seio

İmza/Tarih

31.12.2024

Aydınlatılmış Onam Formu

Sayın Okul Müdürü,

“Ortaokul 7. Sınıf Matematik Dersinde Açık Uçlu Soruların Değerlendirilmesi Yapay Zekâ ve Öğretmen Puanlamalarının Karşılaştırılması” adlı araştırmaya katılmak üzere davet edilmiş bulunuyorsunuz. Bu araştırmada yer almayı kabul etmeden önce, araştırmanın ne amaçla yapılmak istendiğini anlamanız ve kararınızı bu bilgilendirme çerçevesinde özgürce vermeniz gerekmektedir. Aşağıdaki bilgileri lütfen dikkatlice okuyunuz, sorularınız olursa sorunuz ve açık yanıtlar isteyiniz.

Bu araştırmanın amacı sınıf içi ölçmelerde kullanılan açık uçlu maddelerden oluşan bir matematik testinde, Yapay Zeka (YZ) ve öğretmen değerlendirmelerinin etkinliğini ve güvenilirliğini karşılaştırmak; yöntemlerin test istatistikleri ve sonuç raporları arasındaki tutarlılıkları ve farklılıklarını incelemek ve her öğrenciye özelleştirilmiş geri bildirim sağlayan karnelerin hazırlanmasında YZ ve öğretmen değerlendirmesi yöntemlerinin öğrenci başarısı ve motivasyonu üzerindeki etkilerini analiz etmektir. Araştırma için 200 öğrenciden veri toplanması hedeflenmektedir. Araştırma gizlilik esasına dayalı olarak yürütülecektir. Katılımcılardan elde edilen veriler araştırma kapsamı dışında kullanılmayacaktır.

Bu araştırmada yer almak tümüyle sizin isteğinize bağlıdır. Araştırmada yer almayı reddedebilirsiniz ya da başladıktan sonra yarıda bırakabilirsiniz. Bu araştırmanın sonuçları bilimsel amaçlarla kullanılacaktır. Araştırmadan çekilmeniz ya da araştırmacı tarafından araştırmadan çıkarılmanız halinde, sizle ilgili veriler kullanılmayacaktır. Ancak veriler bir kez anonimleştikten sonra araştırmadan çekilmeniz mümkün olmayacaktır. Sizden elde edilen tüm bilgiler gizli tutulacak araştırma yayınlandığında da varsa kimlik bilgilerinizin gizliliği korunacaktır. Ayrıca araştırmada sizden herhangi bir ücret talep edilmeyip, herhangi bir ücret de ödenmeyecektir.

Yukarıda yer alan ve araştırmaya başlanmadan önce gönüllülere verilmesi gereken bilgileri içeren metni okudum (ya da sözlü olarak dinledim). Eksik kaldığını düşündüğüm konularda sorularımı araştırmacılara sordum ve doyurucu yanıtlar aldım. Yazılı ve sözlü olarak tarafıma sunulan tüm açıklamaları ayrıntılarıyla anladığım kanısındayım. Çalışmaya katılmayı isteyip istemediğim konusunda karar vermem için yeterince zaman tanındı. Bu koşullar altında, araştırma kapsamında elde edilen şahsıma ait bilgilerin bilimsel amaçlarla kullanılmasını, gizlilik kurallarına uyulmak kaydıyla sunulmasını ve yayınlanmasını, hiçbir baskı ve zorlama altında kalmaksızın, kendi özgür irademle kabul ettiğimi beyan ederim.

Gönüllünün adı/soyadı

Merve BAŞAR

İmza/Tarih

Araştırmacının adı/soyadı

Serriq Setir Seio

İmza/Tarih

20.12.2024

BENZERLİK BİLDİRİMİ

“AÇIK UÇLU MATEMATİK SORULARININ DEĞERLENDİRİLMESİNDE YAPAY ZEKÂ VE ÖĞRETMEN PUANLAMALARININ KARŞILAŞTIRILMASI” başlıklı tezimin ana bölümü (ön bölüm, kaynaklar ve ekler hariç) Turnitin İntihal Tespit Programı aracılığıyla incelenmiş ve ilgili rapor danışmanım tarafından da kontrol edilmiştir. Kontrol sırasında (1) “Yedi sözcükten daha az olan benzeşmeler” (2) “Kaynaklar” (3) “Doğrudan alıntılar” ve (4) “Lisansüstü tezim ile ilgili yapmış olduğum kendi yayınlarım ve yararlandığım mevzuat metinleri gibi kaynaklar” dışarıda tutulmuştur. Benzerlik kontrolüne ilişkin rapordan elde edilen sonuçlar aşağıda sunulmuştur.

Rapor Tarihi	: 29-Tem-2025 10:31 +03
Gönderim Numarası	: 2722304822
Sayfa Sayısı	: 115
Sözcük Sayısı	: 20596
Karakter Sayısı	: 205233
Benzerlik Oranı	: %6
Savunma Tarihi	: 10.07.2025

Yukarıda belirtilen sonuçları gösteren Turnitin İntihal Tespit Programı’na ilişkin orijinal raporu, sonuçlarda herhangi bir değişiklik yapmaksızın bu beyanım ekinde Enstitüye teslim ettiğimi, tezimin %10’dan fazla benzerlik oranı içerdiğinin, tek bir kaynakla eşleşme oranının ise %2’den fazla olduğunun belirlenmesi durumunda, bundan doğabilecek tüm yasal sorumluluğu kabul ettiğimi bildirir, saygılarımı sunarım.

Öğrencinin Adı Soyadı: Serra Sezin Şener

Tarih: 29.07.2025

İmza:

ÖZGEÇMİŞ

Kişisel Bilgiler

Adı ve Soyadı : Serra Sezin Şener

E-Posta Adresi :

İş Deneyimi :

Unvan	Görev Yeri	Yıl
Memur	Maya Okulları Oran Yerleşkesi	2012 -
Danışman	Türkiye Büyük Millet Meclisi	2008 - 2009

Akademik Bilgiler

Öğrenim Durumu:

Derece	Bölüm/Program	Üniversite	Yıl
74,71	Uluslararası İlişkiler	Ankara Üniversitesi	2003

Yayınlar: