Determining the Role of Artificial Intelligence-Supported Applications in Reliability of Measurement Activities: Awareness and Precautions

Ölçme Faaliyetlerinin Güvenirliğine Yönelik Yapay Zeka Destekli Uygulamaların Rolünün Belirlenmesi: Farkındalıklar ve Önlemler

Servet KILIÇ

Doç. Dr.◆ Ordu Üniversitesi, Bilgisayar Teknolojileri Bölümü ◆ servetkilic@odu.edu.tr ◆ ORCID: 0000-0002-1687-3231

Abstract

This study aims to explore the role of artificial intelligence (AI)-supported applications in enhancing the reliability of online or face-to-face measurement activities. Using a case study design, a qualitative research method, the study gathered the opinions of 48 educators from various disciplines in schools affiliated with the Ministry of National Education and Higher Education. Data were collected using a semi-structured interview form consisting of four open-ended questions. The data were analyzed with the content analysis method and the Cohen's kappa reliability coefficient was determined as k = .78, p < .05. The results revealed that nearly half of the educators were aware of AI-supported applications (such as ChatGPT, Bing, Copilot, etc.) and that some students used these applications to answer questions or prepare homework during the measurement process. In this context, the study identified various precautions that can be implemented to increase the reliability of online or face-to-face measurement activities. These precautions included conducting exams in person whenever possible, using interpretative questions, limiting exam durations, and focusing on practical exams. Additionally, educators can take steps such as recording screens during exams, capturing live images with cameras, and monitoring opened tabs to reduce the impact of AI-supported tools. The current study further discusses the influence of these measurement methods and precautions on increasing the reliability of measurement results.

Keywords: Artificial intelligence, Measurement reliability, ChatGPT

Özet

Bu çalışmanın amacı, çevrimiçi veya yüz yüze yapılan ölçme faaliyetlerinin güvenirliğine yönelik yapay zeka (YZ) destekli uygulamaların rolünün belirlenmesidir. Nitel araştırma yöntemleri arasında yer alan durum çalışması deseni kullanılan çalışma kapsamında, Yükseköğretim'e ve Milli Eğitim Bakanlığı'na bağlı okullarda farklı branşlarda görev yapan 48 eğitimcinin görüşleri alınmıştır. Veri toplama aracı olarak açık uçlu dört sorudan oluşan yarı yapılandırılmış görüşme formu kullanılmıştır. Veriler içerik analizi yöntemiyle analiz edilmiş ve Cohen's kappa güvenirlik katsayısı k = .78, p < .05 olarak belirlenmiştir. Çalışma sonuçlarına göre, eğitimcilerin neredeyse yarısının YZ destekli uygulamaların (ChatGPT, Bing, Copilot vb.) farkında olduğu ve ölçme sürecinde bazı öğrencilerin soruları cevaplamak veya ödevleri hazırlamak amacıyla bu uygulamalardan yararlandığı ortaya konulmuştur. Bu kapsamda çevrimiçi veya yüze yüze yapılan ölçme faaliyetlerinin güvenirliğini artırmaya yönelik eğitimciye ve yazılıma bağlı olarak alınabilecek önlemler belirtilmiştir. Sınavların olabildiğince yüz yüze gerçekleştirilmesi, yoruma dayalı soruların kullanılması, sınav süresinin kısıtlanması ve uygulamaya dayalı sınavların yapılması eğitimcinin alabileceği önlemler arasındadır. Sınav esnasında ekran kaydının tutulması, kamera ile canlı görüntünün alınması ve açılan sekmelerin kontrol edilmesi yazılıma bağlı olarak alınabilecek önlemler arasındadır. Eğitimcilerin benimsedikleri ölçme yöntemlerinin ve belirttikleri önlemlerin ölçme sonuçlarının güvenirliğini artırmaya yönelik etkileri tartışılmıştır.

Anahtar Kelimeler: Yapay zeka, Ölçme güvenirliği, ChatGPT

1. Introduction

Although there are different definitions in the literature on Artificial Intelligence (AI), AI, in its most general definition, is the efforts to transfer some high-level mental skills, such as problem-solving, reasoning, analysis, interpretation, and evaluation- associated with human intelligence to machines through computer programs (Mathew & Mani, 2023). With the rapid development of technology, AI seeks to simplify daily life. By enabling machines or applications to think like humans, AI has become effective in many areas including healthcare, education, commerce, tourism, transportation, communication, and entertainment. Healthcare workers are helped to diagnose and treat diseases faster. Thanks to automatic grading, individualized applications, and adaptable systems, educators contribute to the development of students (Swiecki et al., 2022). Companies can more easily determine customer demands and trends. An AI-supported application can process the preferences and interests of people based on the videos they watch into a database and then present content according to users' experiences. Thus, although the application used is the same, the experience is personalized. Logistics and transportation can be made efficient thanks to traffic management, vehicle routing, and transportation planning. Multi-purpose mobile phones, mobile applications, and virtual assistant assistants work in harmony with AI and take their place in people's lives (Ng et al., 2021; Yang, 2022).

Along with its use across various fields, AI is increasingly being integrated into both face-to-face and online education. In Turkey, AI was highlighted in the 2023 education vision of the Ministry of National Education (MoNE). This report emphasized the effect of AI in enabling machines to behave like humans, in the realization of learning, and in reaching the breaking point of the monopoly of humans intelligence in education. Therefore, it was emphasized that AI has become too important to be ignored and that it can be integrated into the learning process. There are different studies in the literature on the effect of AI-supported applications on learning. Personalized learning experience, adaptive learning environments, automatic evaluation and feedback, student tracking, online education, and accessibility offer important opportunities (Aşık et al., 2023). With the pandemic, the tendency towards online education has increased massively, and in this period, there has been an increase in applications that will ensure the effectiveness of education, and studies have been conducted on the integration of AI into learning environments (Chen et al., 2021; Kong et al., 2021). AI is used to create individualized educational content, offer innovative teaching strategies and methods, analyze data make evaluations, and ensure teacher-student interaction (Chassignol et al., 2018). In order to determine the strengths and weaknesses of students, personalized learning experiences can be offered to students based on the resources they use (written course documents, videos, animations, or interactive applications) and their learning experiences, such as the questions they solve. In this way, teachers can identify students who have learning difficulties in various subjects and make the necessary interventions.

AI-supported applications for education and information acquisition can take various forms. These include personalized learning platforms, intelligent educational assistants, automatic exam and evaluation systems, virtual teaching assistants, smart search engines, scientific research platforms, and chatbots. Chatbots based on technologies such as deep learning, machine learning, contextual meaning, image, and language processing are becoming more widespread today. Chatbots such as ChatGPT, Google Bard, Bing Chat, and Claude include features such as providing information, answering questions, producing content, explaining course materials, performing mathematical calculations, making predictions based on data, writing programs, language translation, producing artistic texts such as poetry and solving problems. ChatGPT, which is widely used among chatbots, can answer questions about verbal and numerical fields, health, guidance, language and translation,

technical and programming, entertainment and culture, daily life, problem-solving, and calculation (OpenAI, 2023). As can be seen, thanks to chat robots, answers can be obtained to questions at different levels, such as knowledge, comprehension, analysis, synthesis, and evaluation in many different areas, from daily life to education. When chatbots are used to obtain information, they can contribute to student learning. However, when these robots are used in the process of answering written (open-ended, multiple choice, fill-in-the-blank, short-answer, etc.) or homework-based assessment questions in face-to-face or online education, they can also cause problems regarding the reliability of the assessment activity. Assessment activities are very important in order to determine the functionality of teaching programs, to reveal the effectiveness of teaching strategies, methods, and techniques, and to evaluate students' weaknesses and strengths. Some problems related to evaluation may arise after measurement activities that are not carried out reliably.

There are studies in the literature that determine the accuracy rate and success of chatbots in exams. Göktaş (2023) aimed to determine what kind of results ChatGPT would produce in the accuracy and verification stages in homework and multiple-choice exams to be held in the tourism field. After the study, it was determined that AI-supported applications such as ChatGPT pose a significant threat to the reliability of exams. It was seen that it was quite successful, especially when it came to multiple-choice questions. In homework questions, some incorrect answers and the fact that they were based on non-existent sources caused ChatGPT's reliability in such questions to be questioned. Göktaş (2023) emphasized the need to protect the integrity of online exams and ensure fair and valid assessments for all students. It was also stated that educators and institutions should be aware of the potential for ChatGPT to facilitate cheating and that it is important to take steps to combat this. Choi et al. (2023) tried to reveal to what extent ChatGPT could be successful in law school exams. More than 95 multiple-choice questions and 12 essay questions were asked to ChatGPT. After the study, it was determined that ChatGPT received a passing grade of "C+" on average, albeit low. The study provided sample recommendations on how chatGPT can help students in legal writing rather than limiting its use in the assessment process (Choi et al., 2023). In their study, Bordt and Luxburg (2023) aimed to determine the success of ChatGPT by subjecting it to an exam within the scope of the subject of "Algorithms and data structures" in the field of computer science at the undergraduate level. After the evaluations, it was determined that ChatGPT passed the average by receiving 20.5 points out of 40 and could be successful in undergraduate exams. The study noted that it was insufficient to conclude that ChatGPT had any understanding of computer science. As can be seen, the studies reveal the level of success of chatbots like ChatGPT in exams. Especially considering the pandemic process, education at other levels, especially in higher education, has been conducted online for several semesters. While online education continues in many universities, homework and project assignments can be given to students in face-to-face courses focused on measurement. Keeping the measurement and evaluation process in the background in online education and focusing more on the course execution process can often cause the issue of reliability to be ignored (Jung, 2005). It is stated that students think of cheating in exams given in online education whenever they have the opportunity (King et al., 2009) and that they can do this idea more and more easily than in face-to-face exams (Hillier, 2014). Considering that similar courses will be conducted online in the future, solutions should be put forward to ensure the reliability of measurement activities against AI-supported applications.

Today, students can benefit from AI-supported applications, and some teacher candidates and educators can especially benefit from chatbots (Çam et al., 2021). However, due to reasons such as the rapid development of technology, the inadequacy of some educators' technological literacy at this point, and the lack of information in educational institutions, they may not be aware of such

applications. It is thought that educators, who have an important place in teaching activities, should benefit from AI-supported applications and be aware of what problems may arise when these applications are used for purposes other than their intended purposes (Popenici & Kerr, 2017). In this context, it is very important to obtain information from educators and to reveal the precautions that can be taken against the unethical uses mentioned above. In this context, this study aims to determine the measurement methods used by educators, to reveal the purposes for which educators and students benefit from AI-supported applications, and to determine the precautions that can be taken to increase the reliability of measurement activities within the framework of these applications.

1.1. Research Questions

1. What are the measurement methods used by educators in the teaching process?

2. What is the awareness of educators regarding AI applications and their usage of these applications?

3. What are the situations in which students benefit from AI applications in the measurement process?

4. What are the precautions that can be taken to increase the reliability of measurement activities?

2. Method

This study was conducted using a case study design, which is one of the qualitative research methods. A case study is an in-depth qualitative research approach that explores specific situations within a defined process, with the goal of identifying and analyzing the underlying reasons behind these situations. (Creswell & Poth, 2016). Case studies are used to define the details that make up an event, to expand the possible situations related to the event, and to evaluate events from various aspects (Patton & Appelbaum, 2003).

2.1. Participants

Educators from various disciplines and positions Higher Education Institutions and the MoNE participated in the study. Due to the diversity of participants and the possibility of reaching a sufficient sample group within the province, the researcher preferred the convenience sampling method. After the decision of the ethics committee was taken, the researcher contacted educators in schools within the borders of the province where he was located with different methods (e-mail, online voice interview, face-to-face interview). Information about the details of the study was shared with the educators. Some educators who were contacted did not accept to participate in the study due to their workload and different excuses. A total of 48 educators, 15 female (31%) and 33 male (69%) from different levels of education participated in the study. Participant characteristics are given in Table 1. The number of educators who have worked for 5 to 10 years is 3 (6%), the number of educators who have worked for 11 to 20 years is 27 (56%), and the number of educators who have worked for 20 or more years is 18 (38%).

Table 1. Characteristics of the Participants

Organisation	Department/Branch	Academic title	f	%
University	Computer and Educational Technologies Education	Assoc. Prof. Dr.	5	10
		Dr. Lecturer	1	
		Lecturer	1	
	Software Engineer	Dr.	1	10
	Mechatronics	Dr. Lecturer	1	
	Social Studies	Dr. Lecturer	1	
MoNE	Turkish	Teacher	7	14
	Mathematics	Teacher	7	14
	Information Technologies	Teacher	5	10
	Turkish Language and Literature	Teacher	4	7
	Science	Teacher	3	6
	Preschool	Teacher	3	6
	Religious Culture and Moral Knowledge	Teacher	2	3
	Technology and Design	Teacher	1	
	Geography	Teacher	1	
	English	Teacher	1	
	Class Teacher	Teacher	1	20
	Guidance	Teacher	1	
	Philosophy	Teacher	1	
	Physical Education	Teacher	1	

2.2. Data Collection Tool

Data were collected through a semi-structured opinion form developed during the research process. A detailed literature review was conducted before preparing the opinion form. Studies on AI-supported applications and chatbots, in particular, were examined. Among these studies, studies on candidate or active educators were focused on. The first draft questions for the opinion form were created based on the results and suggestions of the studies. In order to obtain opinions on the suitability of the questions for the purpose and scope of the study, opinions were obtained from two researchers who are experts in the field and work on computer science and educational technologies. In line with the feedback obtained, the questions were revised again, and the opinion form was finalized. The link to the form created online was shared with the educators. Participants were informed that the data obtained would not be shared on different platforms, that the data would be confidential, and that it could be deleted from the database later on, depending on the participant's request. The following questions were included in the opinion form.

1. What are the measurement methods you usually prefer during the teaching process (online and face-to-face)?

2. Do you use AI-supported applications (Chatgpt, Google Bard, Bing AI, etc.)? If yes, please write down the applications you use and specify the purpose for which you use them.

3. Do you think that students benefit from AI applications during the teaching process (online and face-to-face)? If yes, give examples of how students benefit from the applications.

4. Explain the measures that can be taken to increase the reliability of measurement activities within the framework of AI-supported applications.

2.3. Data Analysis

The data was analyzed and interpreted using the content analysis method. Qualitative research data is analyzed by going through stages including coding, extracting themes, categorizing codes according to themes, and defining and interpreting the findings (Yıldırım & Şimşek, 2013). Content analysis involves categorizing related data in a way that is clear to the reader, organizing it into themes, converting findings into simple numerical data using codes, and interpreting them impartially (Koçak & Özgür, 2006; Selçuk et al., 2014). Participants were coded as P1 and P2 and processed. Descriptive inferences were made to support some findings that emerged after the content analysis, and direct quotes were made from the participants. The opinion form was examined in detail and analyses were conducted by two researchers using Nvivo content analysis software. The separate responses given to the form were combined in a file and transferred to the qualitative analysis software. After coding, Cohen's Kappa reliability coefficient, calculated to determine the agreement between the researchers, was found to be .78 (p< .05). Landis and Koch (1977) stated that this value was a very good level of agreement.

2.4. Ethical Considerations

Ethics committee permission was given by Ordu University Educational Research Ethics Committee on 18/07/2024 with the decision number 2024-128. Participants participated in the research voluntarily.

3. Results

The findings were organized on the sub-problems of the study. Since short-answer statements were used in the questions related to the first two problems, the findings were presented through codes, along with their frequency distributions and percentage rates. Since long-answer statements were used in the questions related to the last two problems, the frequency distributions of the themes and codes were shown through figures, making findings more comprehensible.

3.1. Measurement Methods Used by Educators in the Teaching Process

The aim was to determine the measurement methods used by educators in face-to-face or online training. Educators were able to give multiple answers to this question and did not separate measurement methods according to their use in online or face-to-face classes. The measurement methods that educators generally prefer are presented in Table 2.

Table 2. Measurement Methods Used by Educators

Measurement Method	f	%
Open-Ended Questions	31	42
Multiple Choice Questions	21	28
Homework, Project	8	11
Practical Exam	5	7
In-Class Participation	5	7
Fill in the Blank	3	4

In measurement activities, educators generally prefer open-ended and multiple choice exams (N=70, 70%). In order to support the knowledge and skills acquired in the classroom, some educators may give students homework or projects (N=8, 11%). Although limited, educators measure student

success in some courses through practice-based exams or by taking into account active participation in the class (N=10, 14%).

3.2. Educators Benefit from Artificial Intelligence Applications in the Teaching Process

An attempt was made to reveal educators' awareness of AI and the extent to which they benefit from its applications. The findings obtained for this purpose are presented in Table 3.
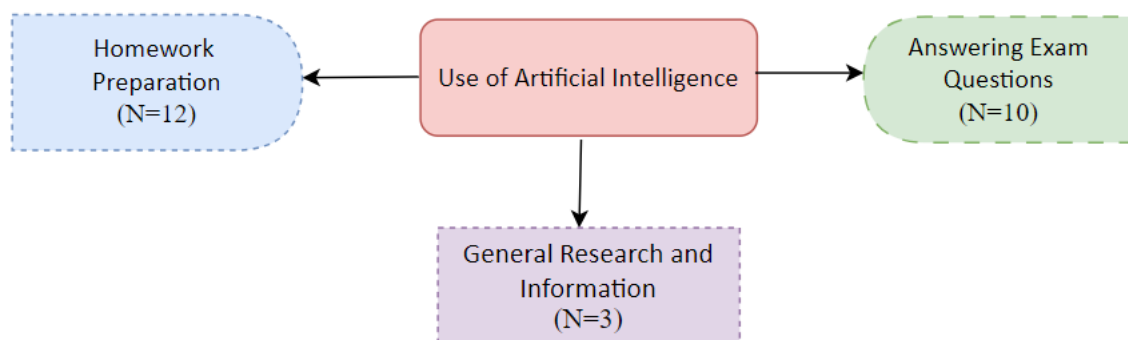
Table 3. Educators' Use of AI-supported Applications

Situation	Applications	f	Purpose	f	%
Yes	ChatGPT	15	Researching	12	25
	Other applications (Gemini, Bing, Copilot etc.)	8	Preparing texts	5	11
			Preparing questions	4	8
			Translating	2	4
No	-	25	-	25	52

Almost half of the educators participating in the study (N=23, 48%) generally use AI-supported applications. Educators who use AI-supported applications mostly prefer the ChatGPT chatbot (N=15, 65%). Educators generally use chatbots for research purposes. Depending on the institution or branch, some educators also use applications for preparing texts, preparing exam questions, or translating.

3.3. Students' Use of AI Applications During the Measurement Process

In the measurement process, an attempt was made to determine the level of students' use of AI-supported applications. In this context, the findings that emerged after the opinions of the educators are shown in Figure 1.

Figure 1. Students' Use of AI-supported Applications

Almost half of the educators who participated in the study (N=22, 46%) stated that students benefit from AI-supported applications in the assessment process. A limited number of students use applications for general research and information acquisition purposes outside of the assessment process (N=3, 6%). When Figure 1 is examined, students use applications to prepare their assignments in face-to-face or online courses. Students mostly prefer the ChatGPT chatbot when preparing assignments. Sample participant statements regarding these findings are given below.

Students use platforms such as ChatGPT, Gamma and Canva, especially when preparing homework and presentations. (P10)

I have witnessed students having ChatGPT prepare a written homework or report. (P11)

I work in the software department so I can say that I am knowledgeable about this subject. Students can have ChatGPT do their homework. (P13)

Yes, I have witnessed students benefiting from AI-supported applications. Students use them for their presentation assignments.(P30)

Since students cannot use any information and communication tools in face-to-face exams, they use applications to answer exam questions, especially in online assessment activities (N=10). Sample participant opinions regarding these findings are presented below.

Students used artificial intelligence tools to cheat in exams during online education. (P11)

I witnessed students answering questions, especially using chatbots, during online education. They copy the question directly to the chatbot and learn the answer, or they have math problems detected through the phone's camera and learn the answer. (P17)

There were students who asked questions to artificial intelligence tools and solved the exam this way. (P8)

3.4. Measures That Can Be Taken to Increase the Reliability of Measurement Activities

An attempt was made to determine the precautions that should be taken to prevent students from benefiting from AI-supported applications during the measurement process and to increase the reliability of measurement activities. The suggestions provided by the educators were grouped under two themes: for the educator (N=22) or for the software (N=10). The codes that emerged depending on the themes are shown in Figure 2.

Figure 2. Suggestions to Increase the Reliability of Measurement Activities
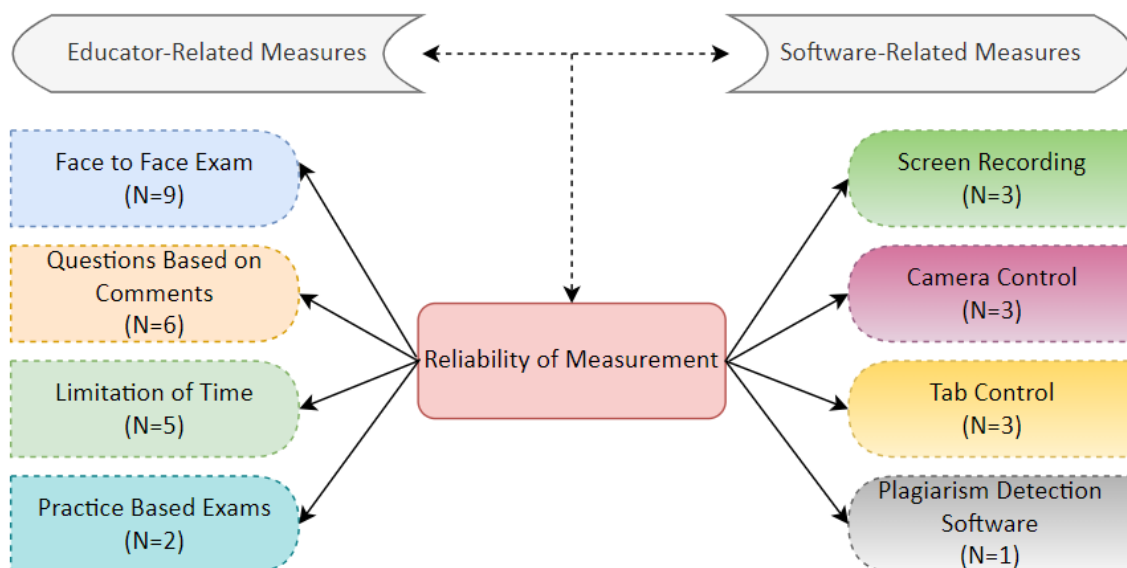
Educators conducting exams face-to-face (N=9) is a serious measure to control students and prevent them from accessing AI applications. Sample participant statements regarding these findings are given below.

Assessment activities can be conducted face-to-face in a way that students cannot use technological devices. (P13)

In online exams conducted outside of face-to-face exams, students will benefit from artificial intelligence. (P39)

Assessment needs to be conducted face-to-face in the school environment. Students should be restricted from using technological devices. (P48)

It was stated that if the questions asked in the exams were arranged based on a higher level of interpretation rather than knowledge level (N=6), students would be able to write their own opinions and thoughts instead of copying information from different sources. Sample participant opinions regarding these findings are presented below.

The reliability of assessment activities can be increased with personalized questions based on interpretation. (P15)

Face-to-face exams should be preferred, and if they are online, internal and interpretation-based assessment should be done. (P24)

Making the questions a little more complex and asking them at the synthesis and evaluation level can also prevent cheating in a way. (P47)

It was stated that the long duration of online exams (N=5) could lead to students obtaining information from different sources. In this regard, the educator with code P20 used the expression "*Exam duration can be limited*". The other educator with code P14 stated the opinion "*Time-limited exams and camera-recorded exams should be conducted".* Some educators stated that conducting exams based on application (N=2) instead of exams based on knowledge on paper would prevent students from directly turning to AI-supported applications. In this regard, the educator with code P47 stated the opinion "*Instead of questions directly measuring knowledge, having questions that are a bit more application-based, project-based could reduce cheating in the assessment process*".

The educators, who also stated that some precautions could be taken depending on the software, emphasized that the tabs opened by the students during the exam could be checked (P7, P33 and P40), screen recordings could be taken (P14, P44 and P46) and controls could be made via camera (P9, P14 and P23). Only one educator stated the following opinion: "*It can be determined with software whether the homework is done by artificial intelligence or not, there are platforms working for this purpose (P2).*"

4. Discussion

This study determined that educators across different institutions and educational levels generally prefer written exams such as open-ended, multiple-choice, and fill-in-the-blank questions, during the measurement process. A previous study on teaching staff has revealed that similar methods are preferred (Borich, 2013; Kılıç, 2021). The reasons for favoring these measurement methods include their alignment with course characteristics, ease of preparation or evaluation, and the convenience they offer to students (Kılıç, 2021). In online measurement activities, these methods are adopted due to reasons such as the characteristics of the exam modules, easy preparation, ease of evaluation, and quick completion (Baran, 2020). These measurement methods, which are preferred by many educators, can make it easier for students to copy information from different sources on the internet, both through search engines and AI-supported chatbots, especially in online education. Some educators measure students' success with studies carried out outside of school by giving homework or project assignments. Although homework contributes to reinforcing students' achievements in the courses, it is thought that students can benefit from these applications in homework given directly for information acquisition or especially for problem solving in numerical courses. Although limited, some educators measure student success in practice-based courses in laboratories based on computer

applications. It can be said that measurement activities where knowledge-based skill development is measured under the supervision of an educator, on computers with restricted internet access, and instead of directly measuring students' knowledge, will be more reliable. Indeed, some educators have made suggestions for conducting application-based exams in this study. It can be said that active participation in class, which encourages students' active participation in the course and has less impact on determining students' success, is a preferable method. Instructors have stated that students' active participation in the course based on performance can be included in the scoring and the reliability of measurement activities can be increased by asking students to defend their assignments verbally (Kılıç, 2021).

Almost half of the educators participating in the study use AI-supported chatbots for research, text preparation, question preparation or translation purposes. With the use of AI technology in the field of education in recent years, educators' awareness of these technologies has been increasing (Çam et al., 2021). Chatbots such as ChatGPT are generally preferred by both educators and students thanks to their features such as easily obtaining information in many different fields, solving complex problems, writing texts and making inferences (Kasneci et al., 2023; Rahman & Watanobe, 2023). It has been stated that ChatGPT has a significant effect on completing courses with valid grades in measurement activities conducted in different fields such as law, health, tourism and engineering (Bordt & Luxburg, 2023; Choi et al., 2023; Göktaş, 2023). It is thought that educators who are aware of AI-supported applications and use these applications will be more sensitive and careful to ensure the reliability of measurement activities. These educators can anticipate that some students will benefit from these technologies in the measurement process and can shape their measurement methods, tools and processes accordingly. Especially in online assessment activities, students can share information during the exam and benefit from resources on the internet thanks to the common chat groups created (Kılıç, 2021; King et al., 2009). The free use and easy access of chatbots such as ChatGPT can make it easier for students to benefit from these technologies. In this study, many educators stated that they witnessed students benefiting from AI-supported applications during the assessment process. Students can use these applications while answering exam questions or preparing their homework. It is thought that students tend to use these applications more in online environments where control is difficult, especially during the pandemic and when it is not possible to come together in face-to-face environments. If students who spend most of their learning process online use AI-supported applications during the assessment process, it can lead them to laziness and negatively affect their knowledge and skill development (Pekmez et al., 2024). This situation can cause students to graduate from schools without being equipped and competent enough. Therefore, it is necessary for educators to be aware of AI-supported applications and the features of these applications. The results obtained in this study reveal that some educators are not aware of this issue and do not benefit from the applications.

Educators have stated that some precautions can be taken regarding the measurement methods, tools and modules used to ensure the reliability of measurement activities. It is recommended that measurement activities be carried out face-to-face as much as possible, questions should be prepared based on interpretation in a way that takes into account students' unique opinions, question answering times should be kept short and application-based exams should be preferred. During the pandemic process, it has not been possible for educators and students to come together face-to-face. Therefore, it does not seem possible to hold exams face-to-face in a process where all educational activities are carried out online. Today, hybrid (face-to-face and online) education is becoming increasingly widespread (Göksel & Adıgüzel, 2024; Kumaş, 2023). In hybrid education,

measurement activities can be carried out face-to-face. Chatbots such as ChatGPT can answer open-ended, multiple-choice, interpretation, inference or problem-solving questions at the knowledge, comprehension, analysis, synthesis and evaluation level (Bordt & Luxburg, 2023; Choi et al., 2023; Göktaş, 2023). Therefore, using measurement tools that include different levels of interpretation-based questions for the purpose of directly obtaining information may negatively affect reliability. ChatGPT can establish relationships between different concepts given to it and make some comments on the desired outcomes. Educators' suggestions for preparing questions based on interpretation may not be sufficient at this point. In online exams, a time limit on the module used may limit students' ability to move more comfortably during the exam and to obtain more information from different sources. It is important to adjust the answering times appropriately, especially in questions prepared at the level of analysis, synthesis or evaluation based on interpretation. On the other hand, Kolomuç and Çelik (2021) stated that there is no relationship between test completion time and test success in online exams held during the Covid-19 period. Therefore, applications such as ChatGPT can answer the questions posed to them very quickly. Imposing a time limit may not completely eliminate the situation where students benefit from these applications. Depending on the software used, some educators have suggested recording the screen during the exam, turning on the camera to capture students' images during the exam, and checking whether different tabs are opened. Kılıç (2021) stated that actively using devices such as cameras and microphones in the exam modules in learning management systems and performing tab checks during the exam will contribute to exam reliability. Again, in the open education high school exams conducted online by the MoNE, the tabs opened by the students during the exam and the mouse movements on different pages were checked. Students were given a certain amount of time while opening different tabs and the students' exams were completed when the time was up (MoNE, 2021). Exam modules structured with AI support can keep some data about the students' behaviors during the measurement. The data obtained from the camera, microphone and screen recording can be evaluated using some algorithms and inferences can be made about the reliability of the measurement. However, students can obtain information from a different computer or communication tool during the measurement. In this case, taking a screenshot of the computer where the measurement is made and checking the tabs may not be a sufficient measure to ensure reliability. Therefore, while it can be said that it is quite difficult to ensure the reliability of measurement activities carried out especially in a result-oriented manner, it is more appropriate to develop creative projects that measure students' skill development based on performance, conduct exams based on in-class participation or application.

5. Conclusions and Recommendations

This study aimed to examine the effects of AI-supported applications on the reliability of measurement activities. To this end, data was obtained from educators working in higher education institutions and the MoNE regarding the measurement methods they use, their use of AI-supported applications, students' use of these applications during the measurement process, and the precautions that can be taken. In the measurement process, educators generally prefer written exams with open-ended, multiple-choice, or partially fill-in-the-blank questions, while rarely using homework or project assignments. Depending on the structure of the course, some educators incorporate practical exams and encourage active participation. Almost half of the educators interviewed use AI-supported applications for different purposes, such as conducting research, preparing texts, creating questions, or translating, and are aware of both the advantages and disadvantages of these applications. Some

students use AI-supported applications to answer exam questions or complete assigned homework during both online or face-to-face assessments.

In particular, in online exams or assignments where educator supervision is limited, leading to potential problems with the reliability of measurement activities. The measurement methods preferred by educators can inadvertently facilitate students' use of chatbots, such as ChatGPT, to acquire information. In order to increase reliability, educators suggested that exams should be conducted face-to-face, focus on practical applications, limit the time given, and include questions based on critical thinking. The fact that chatbots such as ChatGPT have the features of solving problems in a short time, comparing, relating or making comments according to the given instructions may also render the measures that can be taken in this direction inadequate. Especially in online exams, measures such as taking screen recordings, checking the camera or tabs may be insufficient in ensuring reliability. Therefore, instead of implementing measures that restrict the use of AI-supported applications in the measurement process, it may be more effective to adopt assessment methods that will increase the benefits of these applications while promoting student knowledge and skill development. Including process-oriented and performance-based assessment activities, rather than focusing on outcomes, is expected to be more effective in ensuring measurement reliability. It is essential for policymakers in the field of education to share their perspectives and make decisions aimed at increasing awareness about AI-supported applications and ensuring their integration into education. It is deemed important to inform educators working in schools affiliated with Higher Education Institution or the MoNE about the use of AI-supported applications and the need to reconsider the measurement methods in the context of these applications is crucial. The research sample in this study is limited to 48 educators, so expanding the study by reaching more educators from various disciplines in universities and schools in different cities in terms of the generalizability of the results. Additionally, experimental studies involving project-based tasks within the scope of artificial intelligence applications such as chatGPT, can be conducted to explore the contribution of these applications to learning processes.

References

Aşık, F., Yıldız, A., Kılınç, S., Aytekin, N., Adalı, R. & Kurnaz, K. (2023). The impact of artificial intelligence on education. *International Journal of Social and Humanities Sciences Research (JSHSR), 10*(98), 2100–2107. https://doi.org/10.5281/zenodo.8307107

Baran, H. (2020). Açık ve uzaktan eğitimde ölçme ve değerlendirme. [Measurement and evaluation in open and distance education]. *Açıköğretim Uygulamaları ve Araştırmaları Dergisi (AUAd), 6*(1), 28-40. https://dergipark.org.tr/en/pub/auad/issue/55639/761239

Bordt, S., & Luxburg, V. U. (2023). *ChatGPT participates in a computer science exam*. ArXiv preprint arXiv:2303.09461. https://doi.org/10.48550/arXiv.2303.09461

Borich, G. D. (2013). *Effective teaching methods* (8th ed.). Merrill.

Choi, J. H., Hickman, K. E., Monahan, A. B., & Schwarcz, D. (2021). ChatGPT goes to law school. *J. Legal Educ., 71*, 387.

Chassignol, M., Khoroshavin, A., Klimova, A., & Bilyatdinova, A. (2018). Artificial intelligence trends in education: A narrative overview. *Procedia Computer Science, 136*, 16-24. https://doi.org/10.1016/j.procs.2018.08.233

Chen, C. M., Wang, J. Y., & Hsu, L. C. (2021). An interactive test dashboard with diagnosis and feedback mechanisms to facilitate learning performance. *Computers and Education: Artificial Intelligence, 2*, 100015-100024. https://doi.org/10.1016/j.caeai.2021.100015

Creswell, J. W., & Poth, C. N. (2016). *Qualitative inquiry and research design: Choosing among five approaches*. Sage publications.

Çam, M. B., Çelik, N. C., Güntepe, E. T., & Durukan, Ü. G. (2021). Öğretmen adaylarının yapay zekâ teknolojileri ile ilgili farkındalıklarının belirlenmesi. [Determining teacher candidates' awareness of artificial intelligence technologies]. *Mustafa Kemal University Journal of Social Sciences Institute*, *18*(48), 263-285. https://dergipark.org.tr/en/download/article-file/1764677

Göksel, Ş., & Adıgüzel, A. (2024). Hibrit öğrenme modeli üzerine bir meta sentez çalışması: Uluslararası örnekler. [A meta-synthesis study on hybrid learning model: International examples]. *Milli Eğitim Dergisi*, *53*(243), 1655-1698. https://doi.org/10.37669/milliegitim.1252931

Göktaş, L. S. (2023). ChatGPT uzaktan eğitim sınavlarında başarılı olabilir mi? Turizm alanında doğruluk ve doğrulama üzerine bir araştırma. [Can ChatGPT succeed in distance education exams? A research on accuracy and verification in tourism]. *Journal of Tourism & Gastronomy Studies, 11*(2), 892-905. https://doi. org/10.21325/jotags.2023.1224

Hillier, M. (2014). The very idea of e-Exams: Student (pre) conceptions. In B. Hegarty, J. McDonald, & S.-K. Lok (Eds.), *Rhetoric and Reality: Critical perspectives on educational technology. Proceedings of the Australasian Society for Computers in Learning in Tertiary Education* (pp. 77-88). Ascilite. *https://ascilite.org/conferences/dunedin2014/files/fullpapers/91-Hillier.pdf*

Jung, I. S. (2005). Quality assurance survey of mega universities. Perspectives on Distance Education: Lifelong learning and distance higher education (Ed: C. McIntosh). Vancouver, BC: Commonwealth of Learning & Paris: UNESCO, pp. 79-96. http://oasis.col.org/bitstream/handle/11599/70/PSeries_LLLDHE.pdf?sequence=1#page=89

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kılıç, (2021). Uzaktan eğitimde yapılan ölçme faaliyetlerine yönelik kaygılar ve çözüm önerileri: Ordu Üniversitesi örneği. *ERPA International Congresses on Education.* (P. 553-558). Sakarya, Türkiye.

King, C. G., Guyette Jr, R. W., & Piotrowski, C. (2009). Online exams and cheating: An empirical analysis of business students' views. *Journal of Educators Online*, *6*(1), n1. https://files.eric.ed.gov/fulltext/EJ904058.pdf

Koçak, A., & Arun, Ö. (2006). İçerik analizi çalışmalarında örneklem sorunu. [The sampling problem in content analysis studies]. *Journal of Selcuk Communication*, *4*(3), 21-28. https://dergipark.org.tr/en/download/article-file/177956

Kolomuç, A., & Çelik, S. A. (2021). Covid-19 döneminde yapılan online sınavlarda test bitirme süresi ile test başarısı arasındaki ilişkinin incelenmesi. [Investigation of the relationship between test

completion time and test success in online exams during the covid-19 period]. *Journal of Academic Social Science Studies*, *14*(88),1-12. https://doi.org/10.29228/JASSS.53941

Kong, S. C., Cheung, W. M. Y., & Zhang, G. (2021). Evaluation of an artificial intelligence literacy course for university students with diverse study backgrounds. *Computers and Education: Artificial Intelligence, 2*, 100026. https://doi.org/10.1016/j.caeai.2021.100026

Kumaş, A. (2023). Fen bilimleri derslerinde hibrit eğitim kapsamında aktif öğrenme aracı olarak teknolojinin kullanımı. [The use of technology as an active learning tool in the scope of hybrid education in science courses]. *Trakya Journal of Education, 13*(2), 943-961. https://doi.org/10.24315/tred.1100907

Mathew, A. T., & Mani, P. (2023). Strength of Deep learning-based solutions to secure healthcare IoT: A critical review. *The Open Biomedical Engineering Journal*, *17*(1). https://doi.org/10.2174/18741207-v17-e230505-2022-HT28-4371-2

MoNE (2021). *Açık öğretim kurumları 2020-2021 eğitim öğretim yılı 2. dönem online (çevrim içi) sınavları duyurusu*. Date of Access: 1.11.2024. http://aol.meb.gov.tr/meb_iys_dosyalar/2021_03/19151858_AOYK_2020-2021_EgYitim-OYgYretim_YYlY_2._DoYnem_OnlineCYevrim_IYcYi_SYnavlarY_Duyurusu_.pdf

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence, 2*, 100041-100051. https://doi.org/10.1016/j.caeai.2021.100041

OpenAI, (2023). *Introducing ChatGPT*. Date of Access: 1.11.2024. https://openai.com/blog/chatgpt.

Patton, E., & Appelbaum, S. H. (2003). The case for case studies in management research. *Management Research News*, *26*(5), 60-71. https://doi.org/10.1108/01409170310783484

Pekmez, S., Çoban, T. C., Kılıç, M., & Duman, Y. M. (2024). Eğitimde yapay zeka teknolojilerinin kullanımına yönelik öğretmen görüşleri. [Teachers' views on the use of artificial intelligence technologies in education]. *Ulusal Eğitim Dergisi*, *4*(2), 601-619.

Popenici, S. A., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning*, *12*(1), 22. https://link.springer.com/article/10.1186/s41039-017-0062-8

Rahman, M. M., & Watanobe, Y. (2023). ChatGPT for education and research: Opportunities, threats, and strategies. *Applied Sciences*, *13*(9), 5783. https://doi.org/10.3390/app13095783

Selçuk, Z., Palancı, M., Kandemir, M., & Dündar, H. (2014). Eğitim ve bilim dergisinde yayınlanan araştırmaların eğilimleri: İçerik analizi. [Tendencies of the researches published in education and science journal: content analysis]. *Education and Science, 39*(173), 430-453.

Geniş Özet

Giriş

Yapay Zeka (YZ) ile ilgili literatürde farklı tanımlar yer alsa da en genel tanımıyla YZ, insan zekâsı ile işkillendirilen problem çözme, muhakeme, analiz, yorumlama ve değerlendirme gibi birtakım üst düzey zihinsel becerilerin bilgisayar programları aracılığıyla makinelere aktarılması çabalarıdır (Mathew & Mani, 2023). Birçok alanda kullanılmasıyla birlikte YZ, günümüzde hem yüz yüze hem de çevrim içi eğitime katkı sağlamak amacıyla da tercih edilebilmektedir. Pandemiyle birlikte kitlesel olarak çevrim içi eğitime yönelim artmış, bu dönemde özellikle eğitimin etkililiğini sağlayacak uygulamalarda artış görülmüş ve YZ'nin öğrenme ortamlarına entegrasyonuna ilişkin çalışmalar yürütülmüştür (Chen ve diğerleri, 2021; Kong ve diğerleri, 2021). Öğrencilerin güçlü ve zayıf yönlerini belirlemek amacıyla, yararlandığı kaynaklardan (yazılı ders dokümanları, video, animasyon veya etkileşimli uygulamalar) ve çözdükleri sorular gibi öğrenme yaşantılarından yola çıkılarak öğrencilere kişiselleştirilmiş öğrenme deneyimleri sunulabilir.

ChatGPT, Google Bard, Bing Chat ve Claude gibi sohbet robotları; bilgi sağlama, soruları yanıtlama, içerik üretme, ders materyallerini açıklama, matematiksel hesaplamalar yapma, verilere dayalı tahminler yürütme, program yazma, dil çevirisi, şiir gibi sanatsal metinler üretme ve problem çözme gibi özellikleri içermektedir. Bu robotlar yüz yüze veya çevrim içi eğitim sürecinde yazılı (açık uçlu, çoktan seçmeli, boşluk doldurmalı, kısa cevaplı vb.), veya ödeve dayalı olarak uygulanan ölçme sorularının cevaplandırılması sürecinde kullanıldığında, ölçme faaliyetinin güvenirliliğine yönelik sorunları da ortaya çıkarabilir. Gelecekte de benzer şekilde kitlesel olarak derslerin çevrim içi olarak yürütüleceği dikkate alındığında, ölçme faaliyetlerinin YZ destekli uygulamalara karşı güvenirliğinin sağlanması noktasında çözümler ortaya konulmalıdır. Bu kapsamda bu çalışmada, eğitimcilerin kullandığı ölçme yöntemlerini belirlemek, eğitimcilerin ve öğrencilerin YZ destekli uygulamalardan ne amaçla faydalandıklarını ortaya koymak ve bu uygulamalar çerçevesinde ölçme faaliyetlerinin güvenirliğini artırmak amacıyla alınabilecek önlemleri belirlemek amaçlanmıştır. Bu çerçevede aşağıdaki araştırma sorularına yanıt aranmıştır:

1. Eğitimcilerin öğretim sürecinde kullandığı ölçme yöntemleri nelerdir?

2. Eğitimcilerin YZ uygulamalarına yönelik farkındalıkları ve bu uygulamaları kullanım durumları nelerdir?

3. Öğrencilerin ölçme sürecinde YZ uygulamalarından faydalanma durumları nelerdir?

4. Ölçme faaliyetlerinin güvenirliğini artırma yönelik alınabilecek önlemler nelerdir?

Yöntem

Bu çalışma nitel araştırma yöntemleri arasında yer alan durum çalışması deseni kullanılarak yürütülmüştür. Çalışmaya farklı eğitim kademelerinden 15 Kadın (% 31) ve 33 Erkek (% 69) olmak üzere toplam 48 eğitimci katılmıştır. Katılımcı özellikleri ile ilgili tablo aşağıda verilmiştir. 5 ile 10 yıl arasında görev yapan eğitimci sayısı 3 (% 6), 11 ile 20 yıl arası görev yapan eğitimci sayısı 27 (% 56) ve 20 ve daha fazla yıl görev yapan eğitimci sayısı ise 18 (% 38) dir.

Veriler araştırma sürecinde geliştirilen yarı yapılandırılmış görüş alma formu aracılığıyla toplanmıştır. Soruların çalışmanın amacına ve kapsamına uygunluğuna dair görüş almak için, bilgisayar bilimleri ve öğretim teknolojileri üzerinde çalışmalar yürüten alanında uzman iki araştırmacıdan görüş alınmıştır. Uzmanlardan görüş formu için geri bildirimler alınmıştır. Elde edilen veriler, içerik analizi yöntemiyle analiz edilmiş ve yorumlanmıştır. İçerik analizleri sonrası ortaya çıkan birtakım bulguları desteklemek için betimsel olarak çıkarımlarda bulunulmuş ve

katılımcılardan doğrudan alıntılar yapılmıştır. Kodlamalar sonrasında araştırmacılar arasındaki uzlaşıyı belirlemek amacıyla hesaplanan Cohen's Kappa güvenirlik katsayısı .78 olarak bulunmuştur (p< .05). Landis ve Koch (1977) bu değerin çok iyi bir uzlaşı düzeyi olduğunu ifade etmişlerdir. Görüş alma formunda aşağıdaki sorular yer almıştır:

1. Öğretim sürecinde (çevrim içi ve yüz yüze) genellikle tercih ettiğiniz ölçme yöntemleri nelerdir?

2. Yapay Zeka (YZ) destekli uygulamalardan (Chatgpt, Google Bard, Bing AI vb.) faydalanıyor musunuz? Evet ise, kullandığınız uygulamaları yazarak ne amaçla kullandığınızı belirtiniz.

3. Öğretim sürecinde (Çevrim içi ve yüz yüze) ölçme sırasında öğrencilerin YZ uygulamalarından faydalandığını düşünüyor musunuz? Evet ise, öğrencilerin uygulamalardan faydalanma şekillerine örnekler veriniz.

4. YZ destekli uygulamalar çerçevesinde ölçme faaliyetlerinin güvenirliğini artırmak için alınabilecek önlemleri açıklayınız.

Bulgular ve Tartışma

Eğitimciler yapılan ölçme faaliyetlerinde genellikle açık uçlu, çoktan seçmeli ve boşluk doldurma gibi yazılı sınavları tercih etmektedir (N=55, %74). Eğitimcilerin bazıları ödev veya projeler yoluyla öğrencilerin okul dışında yürüttükleri çalışmaları dikkate alarak ders başarısını ölçmektedir (N=8, %11). Sınırlı da olsa eğitimciler bazı derslerde laboratuvar ortamında uygulamaya veya sınıf ortamında gözleme dayalı olarak öğrencilerin başarılarını ölçmektedir. Daha önce öğretim elemanları üzerine yürütülen bir araştırma da benzer yöntemlerin tercih edildiği ortaya konulmuştur (Borich, 2013; Güneş ve diğerleri, 2023; Kılıç, 2021). Bu ölçme yöntemlerin tercih edilmesinin nedenleri arasında; dersin özelliğine bağlı olması, hazırlamanın veya değerlendirmenin kolay olması ve öğrenciye kolaylık sağlaması gösterilmektedir (Kılıç, 2021). Birçok eğitimcinin tercih ettiği bu ölçme yöntemleri, özellikle çevrimiçi eğitimde öğrencilerin internet ortamında gerek arama motorları gerekse de YZ destekli sohbet robotları yoluyla farklı kaynaklardan bilgi kopyalanmasını kolaylaştırabilmektedir.

Çalışmaya katılan eğitimcilerin neredeyse yarısı (N=23, %48) genel olarak YZ destekli uygulamalardan faydalanmaktadır. YZ destekli uygulamalardan faydalanan eğitimciler daha çok ChatGPT sohbet robotunu tercih etmektedir (N=15, %65). Eğitimciler genellikle araştırma yapmak amacıyla sohbet robotlarını kullanmaktadır. Görev yapılan kuruma veya branşa bağlı olarak, bazı eğitimciler metin hazırlamak, sınav sorusu hazırlamak veya çeviri yapmak amacıyla da uygulamalardan yararlanmaktadır. YZ teknolojisinin son yıllarda eğitim alanında kullanılmasıyla birlikte, eğitimcilerin de bu teknolojilere yönelik farkındalıkları giderek artmaktadır *(Çam ve diğerleri, 2021)*. ChatGPT gibi sohbet robotları, birçok farklı alanda kolay bilgi edinme, karmaşık problemleri çözme, metinler yazma ve çıkarımlarda bulunma gibi özellikleri sayesinde gerek eğitimciler gerekse de öğrenciler tarafından genellikle tercih edilmektedir (Kasneci ve diğerleri, 2023; Rahman, & Watanobe, 2023). YZ destekli uygulamalardan haberdar olan ve bu uygulamaları kullanan eğitimcilerin ölçme faaliyetlerinin güvenirliğini sağlamaya yönelik daha hassas ve dikkatli olacağı düşünülmektedir. Bu eğitimciler bazı öğrencilerin ölçme sürecinde bu teknolojilerden faydalanacağını öngörebilir ve bu sayede ölçme yöntemlerini, araçlarını ve süreçlerini ona göre şekillendirebilirler.

Çalışmaya katılan eğitimcilerin neredeyse yarısı (N=22, %46), öğrencilerin ölçme sürecinde YZ destekli uygulamalardan yararlandığını ifade etmiştir. Sınırlı sayıda öğrenci, uygulamaları ölçme süreci dışında genel araştırma ve bilgi edinme amaçlı kullanmaktadır (N=3, %6). Şekil 1 incelendiğinde, yüz yüze veya çevrim içi yapılan derslerde öğrenciler verilen ödevleri hazırlamak için uygulamalardan

yararlanmaktadır. Öğrenciler ödev hazırlarken daha çok ChatGPT sohbet robotunu tercih etmektedir. Bu bulgulara ilişkin örnek katılımcı ifadeleri aşağıda verilmiştir.

> *Öğrenciler, özellikle ödev ve sunum hazırlarken ChatGPT, Gamma ve Canva gibi platformları kullanıyorlar (K10).*

> *Öğrencilerin, yazılı bir ödev ya da rapor istendiğinde bu belgeyi ChatGPT'ye hazırlattıklarına şahit oldum (K11).*

Yüz yüze yapılan sınavlarda öğrenciler herhangi bir bilgi ve iletişim aracı kullanamadıkları için, özellikle çevrim içi yürütülen ölçme faaliyetlerinde, sınav sorularını cevaplamak için uygulamaları kullanmaktadır (N=10). Bu bulgulara ilişkin örnek katılımcı görüşleri aşağıda sunulmuştur.

> *Öğrenciler çevrim içi eğitim sürecinde yapılan sınavlarda yapay zekâ araçlarını sınavlarda kopya çekmek için kullandılar (K11).*

> *Çevrim içi eğitim sürecinde öğrenciler özellikle sohbet robotlarını kullanarak soruları cevaplandırdığına tanık oldum. Doğrudan soruyu sohbet robotuna kopyalayıp cevabını öğreniyorlar veya matematik problemlerini telefonun kamerası yoluyla algılatıp cevabını öğreniyorlar (K17).*

Özellikle çevrim içi yapılan ölçme faaliyetlerinde, öğrenciler oluşturulan ortak sohbet grupları sayesinde sınav anında bilgi paylaşımı yapabilmekte ve internette yer alan kaynaklardan faydalanabilmektedir (Kılıç, 2021; King ve diğerleri, 2009). ChatGPT gibi sohbet robotlarının ücretsiz kullanıma ve kolay erişime sahip olması, öğrencilerin de bu teknolojilerden faydalanmasını kolaylaştırabilmektedir. Öğrenciler sınav sorularını cevaplarken veya ödevlerini hazırlarken bu uygulamaları kullanabilmektedir. Özellikle pandemi sürecinde kontrolün zor olduğu çevrim içi ortamda ve yüz yüze ortamlarda bir araya gelmenin mümkün olmadığı durumlarda öğrencilerin daha fazla bu uygulamalara yönelim gösterdikleri düşünülmektedir.

Eğitimcilerin sınavları yüz yüze yapması (N=9), öğrencileri kontrol etmek ve YZ uygulamalarına erişimini engellemek için ciddi bir önlemdir. Bu bulgulara ilişkin örnek katılımcı ifadeleri aşağıda verilmiştir.

> *Ölçme faaliyetleri öğrencilerin teknolojik cihazları kullanamayacakları şekilde yüz yüze gerçekleştirilebilir (K13).*

> *Yüz yüze sınav harici yapılan online sınavlarda öğrenciler yapay zekadan yararlanacaktır (K39).*

Sınavlarda sorulan sorular bilgi düzeyinden ziyade daha üst seviyede yoruma dayalı olarak düzenlenirse (N=6), öğrencilerin farklı kaynaklardan bilgi kopyalaması yerine kendi görüş ve düşüncelerini yazabileceği belirtilmiştir. Bu bulgulara ilişkin örnek katılımcı görüşleri aşağıda sunulmuştur.

> *Yoruma dayalı kişileştirilmiş sorularla ölçme faaliyetlerinin güvenirliği artırılabilir (K15).*

> *Sınavlar yüz yüze tercih edilmeli, çevrim içi olacaksa da içsel ve yoruma dayalı ölçme yapılmalı (K24).*

Çevrim içi yapılan sınavlarda sürenin uzun olmasının (N=5) öğrencilerin farklı kaynaklardan bilgi edinebilmesine yol açabileceği ifade edilmiştir. Bu konuda K20 kodlu eğitimci *"Sınav süresi kısıtlanabilir"* ifadesini kullanmıştır. Diğer K14 kodlu eğitimci ise *"Zaman sınırlamalı sınavlar ve kamera kayıtlı sınav yapılması"* şeklinde görüş bildirmiştir. Bazı eğitimciler kâğıt üzerinde bilgiye dayalı sınavlar yerine, uygulamaya dayalı sınavlar yapılmasının (N=2), öğrencilerin doğrudan YZ destekli uygulamalara yönelmesini engelleyeceğini ifade etmişlerdir. Bu konuda K47 kodlu eğitimci, *"Soruların doğrudan bilgi ölçen düzeyde sorular olması yerine, biraz daha uygulamaya dayalı, projeye dayalı olması ölçme sürecinde kopyayı azaltabilir"* şeklinde görüş bildirmiştir.

Yazılıma bağlı olarak bazı tedbirler alınabileceğini de ifade eden eğitimciler, sınav esnasında öğrencilerin açtıkları sekmelerin kontrol edilebileceğini *(K7, K33 ve K40)*, ekran kayıtlarının alınabileceğini *(K14, K44 ve K46)* ve kamera yoluyla kontrollerinin yapılabileceğini *(K9, K14 ve K23)* vurgulamışlardır. Sadece bir eğitimci, *"Ödevlerin yapay zekâ tarafından yapılıp yapılmadığı yazılımlarla tespit edilebilir bu amaçla çalışan platformlar vardır (K2)"* şeklinde görüş bildirmiştir.

Günümüzde hibrit (yüz yüze ve çevrim içi) eğitimler giderek yaygınlaşmaktadır (Göksel & Adıgüzel, 2024; Kumaş, 2023). Hibrit eğitimlerde ölçme faaliyetleri yüz yüze gerçekleştirilebilir. ChatGPT kendisine verilen farklı kavramlar arasında ilişki kurabilir ve istenen çıktılara yönelik birtakım yorumlarda bulunabilir. Eğitimcilerin yoruma dayalı sorular hazırlanması ile ilgili önerileri bu noktada yeterli olmayabilir. Çevrim içi olarak yürütülen sınavlarda kullanılan modül üzerinde süre sınırlaması yapılması öğrencilerin sınav esnasında daha rahat hareket etmesini ve farklı kaynaklardan daha fazla bilgi edinimini sınırlandırabilir. Kılıç (2021), öğrenme yönetim sistemlerinde yer alan sınav modüllerinin kamera ve mikrofon gibi aygıtları aktif olarak kullanmasının ve sınav esnasında sekme kontrollerinin yapılmasının sınav güvenirliğine katkı sağlayacağını belirtmiştir. Yine MEB çevrimiçi yürüttüğü açık öğretim lise sınavlarında, öğrencilerin sınav anında açtığı sekmeleri ve farenin farklı sayfalarda gezinme hareketlerini kontrol etmiştir.

Sonuçlar ve Öneriler

Bu çalışma, YZ destekli uygulamaların ölçme faaliyetlerinin güvenirliğine etkisini incelemeyi amaçlamıştır. Bu kapsamda yükseköğretim kurumu ve MEB'de görev yapan eğitimcilerden, kullanılan ölçme yöntemleri, YZ destekli uygulamaları kullanma durumları, öğrencilerin ölçme sürecinde uygulamalardan faydalanma durumları ve alınabilecek önlemler hakkında bilgiler alınmıştır. Eğitimciler ölçme sürecinde, yazılı sınavlarda genellikle açık uçlu, çoktan seçmeli veya kısmen boşluk doldurmalı soruları tercih ederken, az da olsa ödev veya proje verme yöntemini benimsemektedir. Dersin yapısına göre bazı eğitimciler sınavları uygulamalı yaparken, ders içi aktif katılımı da ölçmeye dahil etmektedir. Görüşme yapılan eğitimcilerin neredeyse yarısı, YZ destekli uygulamaları farklı amaçlar (araştırma yapma, metin hazırlama, soru hazırlama veya çeviri yapma) için kullanmakta ve bu uygulamaların sağlayacağı avantaj veya dezavantajların farkında olmaktadır. Bazı öğrenciler çevrim içi veya yüz yüze gerçekleştirilen ölçme sırasında sınav sorularını cevaplandırma veya verilen ödevleri hazırlamak amacıyla YZ destekli uygulamalardan yararlanmaktadır.

Özellikle eğitimci denetiminin zorlaştığı çevrim içi sınavlarda veya ödevlerde ölçme etkinliklerinin güvenirliğine yönelik sorunlar ortaya çıkmaktadır. Eğitimcilerin tercih ettikleri ölçme yöntemleri, öğrencilerin ChatGPT gibi sohbet robotlarından bilgi edinimini kolaylaştırdığı söylenebilir. Eğitimciler güvenirliğin artırılması amacıyla; sınavların yüz yüze ve uygulamaya yönelik yapılmasını, verilen sürelerin kısıtlanarak yoruma dayalı soruların sorulması gerektiğini belirtmişlerdir. ChatGPT gibi sohbet robotlarının kısa sürede problemleri çözme, karşılaştırma, ilişkilendirme veya verilen yönergelere göre yorumlama yapabilme özelliklerine sahip olması, bu yönde alınabilecek önlemleri de yetersiz kılabilmektedir. Özellikle çevrim içi yapılan sınavlarda ekran kaydının alınması, kamera veya sekme kontrolünün yapılması gibi ortaya konulan önlemlerde güvenirliğin sağlanmasında yetersiz kalabilmektedir. Dolayısıyla ölçme sürecinde YZ destekli uygulamaların kullanımlarını kısıtlayacak önlemler almak yerine, bu uygulamaların avantajlarından faydalanarak öğrencilerin bilgi ve beceri gelişimlerini artıracak ölçme yöntemlerini tercih etmenin daha uygun olacağı düşünülmektedir. Sonuç odaklı olarak yapılan ölçme etkinliklerinden ziyade, süreç odaklı ve performansa dayalı ölçme etkinliklerine yer verilmesinin güvenirliği sağlamada daha etkili olacağı ön görülmektedir. Eğitim alanında çalışmalar yürüten politika yapıcıların YZ destekli uygulamalar ile ilgili farkındalıkları artırmak

ve eğitime entegrasyonunu sağlayabilmek amacıyla birtakım görüşler belirtmesi ve kararlar alması uygun görülmektedir. Yükseköğretim Kurumu'na veya MEB'e bağlı okullarda görev yapan eğitimcilere YZ destekli uygulamaların kullanımına ve kullanılacak ölçme yöntemlerinin bu uygulamalar çerçevesinde tekrar ele alınması gerektiğine ilişkin bilgilendirmeler yapılması önemli görülmektedir.

Yayın Etiği Beyanı

Destek ve Teşekkür

Çatışma Beyanı

Araştırmanın yazarları olarak herhangi bir çıkar/çatışma beyanımız olmadığını ifade ederiz.