Sakarya University Journal of Education

SAKARYA UNIVERSITY
JOURNAL OF EDUCATION

SAKARYA
ÜNİVERSİTESİ

Özkan Ergene
Department of Mathematics and Science Education
Sakarya University
Sakarya - Türkiye
ozkanergene@sakarya.edu.tr

Özlem Canan Güngören
Department of Computer Education and Instructional
Technology
Sakarya University
Sakarya - Türkiye
ocanan@sakarya.edu.tr

Gülden Kaya Uyanık
Department of Educational Sciences
Sakarya University
Sakarya - Türkiye
guldenk@sakarya.edu.tr

Sinem Tarhan
Department of Educational Sciences
Bartın University
Bartın - Türkiye
snmtarhan@bartin.edu.tr

Onur İşbulan
Department of Computer and Instructional
Technologies
Sakarya University
Sakarya - Türkiye
oisbulan@sakarya.edu.tr

Zeynep Demirtaş
Department of Educational Sciences
Sakarya University
Sakarya - Türkiye
zeynept@sakarya.edu.tr

Hıdır Karaduman
Department of Educational Sciences
Anadolu University
Eskişehir - Türkiye
hidirk@anadolu.edu.tr

Merve Savaşcı
Department of Foreign Languages Education
Sakarya University
Sakarya - Türkiye
msavasci@sakarya.edu.tr

Ebru Eskici
Department of Educational Sciences
Bolu Abant İzzet Baysal University
Bolu - Türkiye
eebrdic@gmail.com

Ali Orhan
School of Foreign Languages
Zonguldak Bülent Ecevit University
Zonguldak - Türkiye
ali_orh_an@hotmail.com

Statistics Editors

Ebru Eskici
Department of Educational Sciences
Abant İzzet Baysal University
Bolu - Türkiye
eebrdic@gmail.com

Gülden Kaya Uyanık
Department of Educational Sciences
Sakarya University
Sakarya-Türkiye
guldenk@sakarya.edu.tr

Süleyman Demir
Department of Educational Sciences
Sakarya University
Sakarya - Türkiye
suleymand@sakarya.edu.tr

Levent Ertuna
Department of Educational Sciences
Sakarya University
Sakarya - Türkiye
leventertuna@sakarya.edu.tr

Ethics Editor

Ezgi İnal
Department of Educational Sciences
İstanbul Aydın University
İstanbul - Türkiye
ezgiinal1@aydin.edu.tr

English Language Editor

Merve Savaşcı
Department of Foreign Languages Education
Sakarya University
Sakarya - Türkiye
msavasci@sakarya.edu.tr

Layout Editor

Mehmet Emin Çolak
Coordinatorship of Scientific Journals
Sakarya University
Sakarya - Türkiye
mehmetcolak@sakarya.edu.tr

Yakup Beriş
Coordinatorship of Scientific Journals
Sakarya University
Sakarya - Türkiye
yakupberis@sakarya.edu.tr

Indexing

ULAKBİM TR DİZİN

EBSCOhost

ERIHPLUS
EUROPEAN REFERENCE INDEX FOR THE
HUMANITIES AND SOCIAL SCIENCES

MIAR
Matriz de Información para el
Análisis de Revistas

DOAJ

Since 1898 W WILSON

EBSCO
The Belt and Road Initiative
Reference Source

SHERPA/R●MEO

Contents

Research Article

Working with Young Children with Special Needs in Earthquake Zone: Experiences of Special Education Teachers

Emine Ayyıldız[1*]
Beyza Elif Çiçen[2]
Funda Ün[3]
Hikmet Taha Acartürk[4]
Nisa Demir[5]

[1] Istanbul Medeniyet University, İstanbul, Türkiye, emine.ayyildiz@medeniyet.edu.tr,
[2] Teacher Harun Resit Primary School, İstanbul, Türkiye, cicenelif@gmail.com
[3] Mustafa Karaşahin Primary School, İstanbul, Türkiye, fundaun0265@gmail.com
[4] Sıdıka Doğruöz Special Education Application School, İstanbul, Türkiye, hikmettahaacarturk@gmail.com
[5] New Goal Special Education and Rehabilitation Center, İstanbul, Türkiye, nisad493@gmail.com

*Corresponding author

Abstract: Two devastating earthquakes occurred in the southeast of Türkiye on February 6, 2023. Millions of people were damaged in these Kahramanmaraş-centered earthquakes, which seriously affected 11 provinces. It is known that children and individuals with special needs were most affected by such traumatic events, and young children with special needs are in the group at the highest risk in this respect. In this study, in which we employed qualitative phenomenological approach, we aimed to reveal the experiences of special education teachers who worked or are currently working in the earthquake zone with young children with special needs. In this study, we conducted semi-structured interviews with 12 teachers determined by purposeful sampling and we analyzed the data inductively. The findings demonstrate that the participant teachers provided services not only to young children with special needs but also to all children in need. It is also noted that there were deficiencies in planning for working with children with special needs, there were difficulties in initiating and maintaining educational activities, and teachers' experiences of working in the region had positive effects on them. Since the physical and psychological difficulties of the earthquakes and the need for support in the region will continue for a long time, it is essential to provide training on what to do in disasters to teachers who are currently on duty and those who will be assigned later. The necessity of providing psychological support to teachers during active work in the region and upon their return should also be underlined.

Keywords: Natural Disaster, Earthquake, Young Children with Special Needs, Special Education Teachers, Phenomenological Approach

1. Introduction

Two devastating earthquakes occurred in the southeast of Türkiye on February 6, 2023, at 04.17 and 13.24, respectively (Disaster and Emergency Management Presidency [AFAD], 2023). More than 50 thousand people lost their lives, hundreds of thousands of buildings were destroyed, and 14 million people were affected as a result of these Kahramanmaraş-centered earthquakes, which affected 11 provinces in the region (Anadolu Agency [AA], 2023; Presidency of the Republic of Turkey, Presidency of Strategy and Budget [PoSB], 2023). In addition to being destructive, the 'February 6 earthquakes', can also be seen as the most damaging earthquakes in the country's history due to their devastating impact on the economy and human life. A lot of lives shattered and millions of people's livelihoods destroyed in one day. So, this was really the disaster of the century for Turkiye. (AFAD, 2023) International Federation of Red Cross and Red Crescent Societies [IFRC], 2024).

As in most natural disasters, it is impossible to know the exact time when an earthquake will occur (Indriasari et al., 2018; İşçi, 2008), and individuals and children with special needs are harmed the most by its possible effects. Most of the individuals with special needs move slowly and this is one of the usual reasons for the risk. In addition, the type of disability with the co-current conditions of the individual can also increase this risk (Hashemi, 2018). For instance, while it is thought that individuals with visual impairments can be aware of natural disasters and take the necessary precautions, if these individuals have co-current intellectual disabilities, it will be difficult for them to understand natural disasters and apply the necessary safety skills (Alexander et al., 2012). Young children with special needs constitute the group at the highest risk (Williams et al., 2008). Not having safety skills such as being able to create a life triangle during an earthquake, moving away from risky places, and managing the crisis makes young children with special needs more vulnerable to earthquakes (Aslangiri & Açkalın, 2019).

Earthquakes are considered important traumatic events in human life because they affect life suddenly in a very negative way. It is known that the impacts of these traumatic events can be long-lasting,

particularly for the children (Kurt & Gülbahçe, 2019). The majority of individuals with special needs may be ignored in natural disasters, which causes serious harm or death of these individuals. It has been reported that the probability of children with special needs having a new disability or dying as a result of the earthquake is considerably higher than that of their peers with typical development (TD) (World Federation Occupational Therapy [WFOT], 2007). One reason for the high risk of new disabilities or death is related to the difficulties in reaching individuals with special needs in search and rescue efforts. White's (2006) study clearly showed how these individuals were forgotten and could not be saved during the 2005 Hurricane Katrina. Other studies also indicate that children with special needs and their families who have experienced natural disasters often suffer due to the lack of planning for themselves (Bilik & Akdağ, 2023; Hemingway & Priestley 2006; Masten & Osofsky, 2010; Yarımkaya & Bakkaloğlu, 2024).

In individuals with special needs, adaptation problems such as significant increase in behavioral problems and significant decrease in displaying social communication skills and daily living skills can be observed along with the earthquake disaster. Behavioral problems observed in these individuals can continue to affect them even months later (Valenti et al., 2012). The re-emergence of previously eliminated behavioral problems, the exacerbation of the child's reactions to the events, and the emergence of new problem behaviors are possible changes that may be observed (Söğüt & Kaya, 2024). Since there is a correlation between children's self-regulation skills and social interaction skills, particularly in early childhood, the quick return of young children to school and social settings after a disaster significantly affects their return to normal life (Masten, 2007). In every emergency, schooling is one of the most important requirements for society to stand on its feet, to return to normal life, and to build a new life. Socialization in schools after trauma also plays an important role in the self-regulation and adaptive behavior of young children with special needs (Ager et al., 2010).

Since teachers spend a significant time of the day with children in schools, their support in ensuring that students' lives return to normal and providing a safe and supportive environment after extraordinary situations is essential (Akman & Gündoğdu, 2002). At this point, teachers play an important role to the same extent as parents in helping children adapt to social life again. A study showed that children with special needs stated that they felt safer with their special education (SE) teachers in case of any natural disaster (Ronoh et al., 2015). On the other hand, according to an another study which focused on teachers who support children suffering from traumatic stress, participant teachers noted that they faced difficulties in supporting students in their daily school lives following trauma. Teachers revealed that they did not well equipped to support their students since they had no trauma training and they experienced difficulties engaging their students emotionally (Alisic et al., 2012).

To minimize the traces of the terrible earthquakes, several projects and initiation which provided financial and educational support were launched. Emergency Appeal for Turkey provided support to the immediate and/or long-term needs of people in the earthquake zone, such as food, shelter and livelihoods, first aid and community health, mental health and psychosocial support (IFRC, 2023). The Ministry of National Education started sending volunteer teachers to the region within the days after the earthquake and immediately initiated education and psychosocial support processes (Ministry of National Education [MoNE], 2023a). Volunteer SE teachers also took an active role in these processes and took part in the education of children with special needs, who are considered the most vulnerable segment in earthquakes and all other natural disasters. The General Directorate of Special Education and Guidance Services has prepared an activity book to support teachers' practices in terms of educational efforts and especially reinforce the teachers' psychosocial support activities in the region. This book includes various activities that aim to ensure that all children in the region, especially children with special needs, are affected from the incidents in the region as little as possible, spend their time effectively and their transition and adaptation processes are facilitated (MoNE, 2023b). Other bodies of

the Turkish government initiated some efforts too. For instance, Disability Rights National Action Plan is enured by Ministry of Family, Labour and and Social Services (2023). Within the scope of this action plan, 110 volunteer sign language interpreters served the individuals with hearing impairment in the region. Koç and Yalçın (2023) also pointed out that Ministry of Family, Labour and and Social Services provided care and shelter for the elderly individuals and the individuals with special needs right after the earthquakes. In addition to these, various projects were carried out after the earthquake in order to provide educational and psychosocial support for children with special needs and their families and to ensure that students with special needs will be prepared for a possible earthquake in future (Gazi University, 2023; Foundation for the Raising and Protection of Children with Intellectual Disabilities [ZİÇEV], n.d.).

Some recent studies conducted with the teachers from different grades and/or disciplines at the region. In their study, Arıcı et al. (2023) investigated the impact of the earthquakes on educational activities in Kahramanmaraş province by interviewing 42 teachers who witnessed the earthquakes and continued to teach in the region afterwards. The interviews were conducted with these teachers to assess the difficulties they encountered after the earthquakes. The results showed that issues about education continuity, student participation, security and health conditions, psychological and emotional problems, and management and logistical challenges were emerged. Also, additional efforts were put by the teachers in order to find solutions to students' psychosocial problems. The study is also suggested that students need a kind of post-earthquake education which should include topics such as adjusting to a new life, psychological support, preparedness training for the unexpected events, disaster awareness, values education, and social awareness. Teachers also emphasized the importance of preparing education more inclusive and efficient (Arıcı et al., 2023). Similarly, Polat and Sarıçam conducted a case study with 25 teachers in Hatay province to explore the current issues they experienced as education, nutrition, accomodation, transportation and other environmental, and psychosocial factors. It is quite siginificant that all participant teachers indicated a need for comprehensive psychological support. However, they also emphasized the adverse effects of delayed psychological support. The results showed that accomodation problems of the students and especially school and classroom equipment issues have not only affected students' academic performance but also increased the challenges faced by teachers in the education process (Polat & Sarıçam, 2024). Since the teachers who were and/or are in earthquake region noted some significant issues about supporting their students in terms of educational, social and emotional aspects, investigating the experiences of SE teachers specifically are crucial. In the present study, we aimed to examine the experiences of SE teachers to provide services in the region affected by the Kahramanmaraş-centered earthquakes that occurred on February 6, 2023, and who worked or are currently working especially with young children with special needs.

2. Method

2.1. Research methodology

Qualitative research which is the most suitable method for addressing a research problem when not knowing the specific factors (Creswell, 2012) aims to reveal events or phenomena in a holistic and realistic way within a naturalistic setting (Morgan, 1996). The main feature of a qualitative research is to examine the subject in a comprehensive manner and understand it in depth (Creswell, 2012). Phenomenological approach frequently used in qualitative research intents to explain the meaning of the experiences of the individuals. In other words, phenomenology studies refer to all of the lived experiences (Giorgi, 1997) and focus on understanding participants' experiences with regard to an event or phenomenon and how they explain the meaning of these experiences (Van Manen et al., 2007). The phenomenology approach defines the true nature of the phenomenon by aiming to reveal the basic structure of various experiences through individuals who experience and can reflect on the phenomenon which the research focuses (Merriam, 2013). In the present research, we wanted to

examine the experiences of SE teachers who worked or are currently working in the earthquake zone in working with young children with special needs using phenomenological approach.

2.2. Participants

The participants of this study consist of 12 SE teachers who have worked or are currently working in the provinces affected by the Kahramanmaraş-centered earthquakes. We determined the participants purposefully by criterion and snowball sampling. Criterion sampling aims to achieve a specific quality or characteristic that is desired to be achieved (Kothari, 2004). In snowball sampling, it is possible to reach other individuals who can contribute more to the study in question through one or more participants who have been determined and reached in line with the study's purpose (Patton, 2002). So before we started our research we determined the criteria as follows: a. being a SE teacher, b. having worked or currently working in the affected regions by the current earthquakes, and c. having provided or currently providing educational services to young children (between the ages of 0-8) with special needs. The second and the fourth authors are currently working in SE schools as SE teachers. So, each of them reached out one interviewee through their school environment. After reaching these two teachers meeting the aforesaid criteria, we reached other participants through these interviewees and invited them to participate in the study. The table below contains information about the participants:

Table 1

Information About the Participants

Participant	Gender	Age	City where the participant worked in the region	Date when the participant went to the region	Duration of stay in the region (days)	Is the participant currently working in the region?
P1	M	24	Malatya	1st 26.02.2023	6	No
			Gaziantep	2nd 25.03.2023	7	
P2	F	24	Hatay	03.03.2023	8	No
P3	M	24	Hatay	03.03.2023	8	No
P4	F	24	Gaziantep	26.03.2023	5	No
P5	M	24	Malatya	1st 26.02.2023	6	No
			Gaziantep	2nd 25.03.2023	7	
P6	F	24	Kahramanmaraş	19.03.2023	5	No
P7	M	24	Malatya	26.02.2023	7	No
P8	F	23	Gaziantep	26.03.2023	7	No
P9	M	23	Hatay	03.03.2023	9	No
P10	M	27	Hatay	01.03.2023	17	No
P11	M	26	Hatay	1st 08.02.2023	7	Yes
			Hatay	2nd 18.05.2023	-	
P12	M	25	Hatay	1st 09.04.2023	7	Yes
			Hatay	2nd 27.05.2023	-	

Note: P1, P5, P11 and P12 went to the region twice. 1st and 2nd time of their arrivals indicated at the table. P11's first arrival to the region was only for search and rescue operations.

2.3. Data collection

We conducted semi-structured interviews (Patton, 2002) with the participants to determine the experiences of SE teachers in working with young children with special needs in the earthquake zone. After verbally explaining the purpose and scope of the study to each potential participant, we sent the informed consent form and set a time to interview individuals who decided to take part in the study. Besides the questions in the interview form, the end of each interview, we asked the participants whether there was anything they would like to add to be ensure that the participants reviewed the entire interview. We conducted all the interviews online with an application used for video conference calls. We audio-recorded all interviews and transcribed verbatim them later.

2.4. Data analysis

We employed the inductive analysis approach (Thomas, 2006) to examine participants' experiences related to our study's purpose in depth by systematically examining the raw data we collected from the interviews. In inductive analysis, the researcher presents the results by coding, categorizing, describing the raw data and reaching the main framework, unlike deductive analysis conducted to test an assumption or theory (Thomas, 2006). In this sense, inductive analysis presents a structure that is easily applicable and increases the trustworthiness of the results achieved with its systematic structure (Strauss & Corbin, 1998). First we all read the transcribed files several times and took some notes. Then, we came together and perform open coding by selecting the four interviews that we thought contained the richest data. During the open coding, we marked the participants' excerpts which were related to our research and briefly wrote down the main idea or concept they contained. Then we divided the workload for the other files and completed the coding process. After that, we gathered again and reviewed our codes. We resolved the disagreements if there was any by discussing on it. We created nine categories (subordinate themes) by grouping the codes and we eventually combined subordinate themes into four superordinate themes. After the peer debriefing process, the themes were determined.

2.5. Trustworthiness and research ethics

Trustworthiness of the study was ensured by collaborative work (Gershon, 2009) and peer debriefing (Creswell & Miller, 2000) processes. We worked for the study collaboratively every step of the way. Additionally, we prepared a file containing excerpts of the participants that rationalized the subordinate and superordinate themes we reached and we asked for opinions from two experts working in the field of early childhood SE and also have experiences on conducting qualitative research. The ethics committee document for this study was obtained from the Istanbul Medeniyet University Educational Sciences Ethics Committee with the decision numbered 2023/04-09 dated 08.05.2023.

3. Findings

The results of the present study indicate four superordinate and nine subordinate themes. Although we specifically focused on young children with special needs since the main purpose of our study was to examine the experiences of SE teachers in working especially with such children, the participants mentioned their experiences with all children with special needs in general. They also expressed their experiences with young children through children with special needs and children with TD.

Figure 1

Superordinate and Subordinate Themes
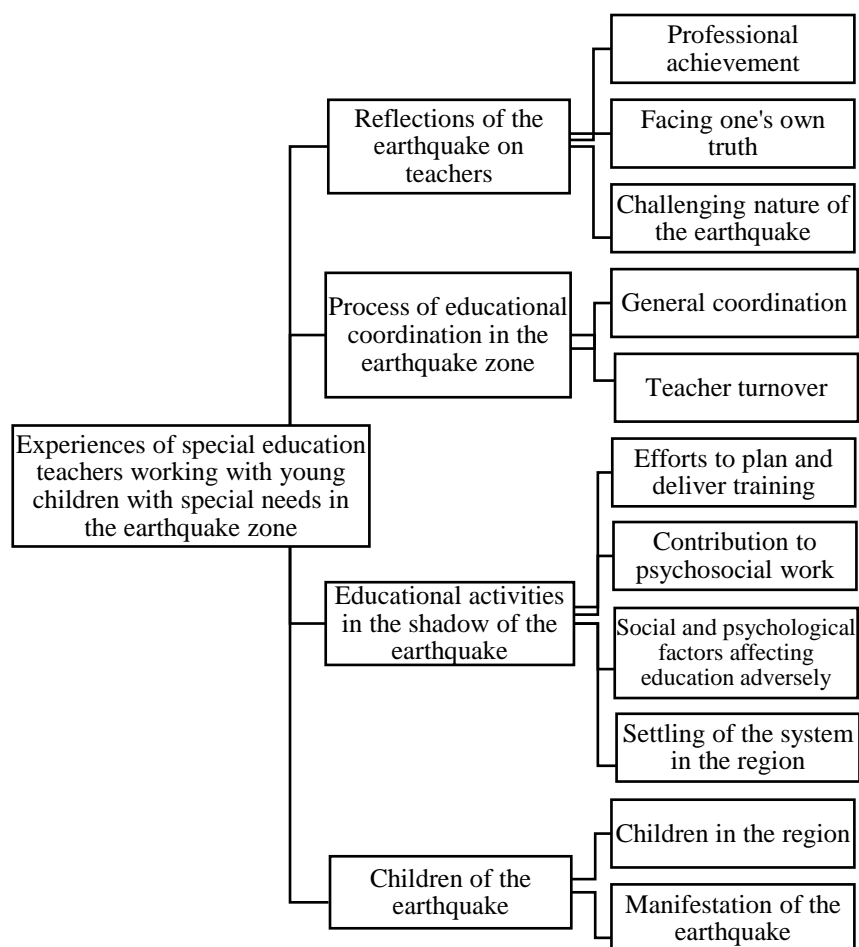
3.1. Reflections of the earthquake on teachers

The findings clearly show that the earthquake had some reflections on the participants concerning their professional achievement and facing their own truth, both personally and professionally. The participants also stressed the effects of the challenging nature of the experiences. Concerning professional achievement, they mentioned how willing they were to go to the region and talked about the contribution of their experience in working with different children, positive effects of going to the region and the experiences they gained. At this point, P4 expressed her experience as, *"I think I was affected in a good way. You know, we saw those students as well in this way. Seeing them and gaining different experiences affected me positively in professional terms."* P1 who went to the region more than once, described his experience of working with different children in following way,

> *"I met an individual with autism… Another child actually had a learning disability… He was between the ages of 6-7… There was also a young age group. There were 17-18-years-old students. There were also a lot of Syrians and Afghans in the region where I was present, that is 50%, maybe more. We were making announcements in Arabic as well as Turkish."*

This same participant noted his experience of how the process was coordinated when he came to the region for the first time as,

> *"Frankly, it was a great experience for me to go there twice. It also benefited me greatly in professional terms. You know, the next time we encounter such a disaster, God forbid! It actually taught us what we could do and how we could approach it. For example, we coordinated and arranged the situation at first. We set up the tents, did scanning, etc. It*

provided us with knowledge and experience about how we can manage the process if we encounter such a disaster again."

P2 emphasized the general positive impact of working in the earthquake zone on teachers by saying the following, *"It was a very good experience. Just think, I am 24 years old, and it makes me very proud to have such an experience."* One of the other important findings that SE teachers questioned themselves personally and professionally while expressing their feelings and thoughts about the earthquake zone. P10 who has been a teacher for six years and worked in the earthquake zone for 17 days said,

"I felt a sense of professional satisfaction. This is the first time I've experienced such satisfaction. I have been teaching for six years, and this is the first time I have been professionally satisfied. In fact, before this I was asking myself 'Did I choose the wrong profession?'. Now it really made me say, 'Being a teacher is in my soul.' I realized that I chose the right profession and was doing a very meaningful job."

P6 who encountered a different situation than she thought when she went to the region, expressed this situation with the following words,

"You know I went there with great motivation. I was saying, 'I'm going to do this and that. I'm strong'. But when I arrived there, I saw what a limited being I was and how limited the areas of my profession actually were."

In addition to the negative emotions experienced by the participants, such as sadness, shock, and inadequacy, the fact that their emotional states were affected by the challenging situations clearly showed the challenging nature of the earthquake. While talking about his work at the beginning with the education of students with special needs, P7 expressed the challenges of the process by saying, *"It was a tiring process, especially to find and screen students."* P12 who is continuing to work in the region, expressed the sadness he experienced with one of the students with as follows,

"I have a student, four years old. He has autism. He lost his entire family. He only has his elder sister and her husband. They rescued him from the wreckage and are looking after him. All his other family members are dead. There are only two of them left and it makes me very sad every time I see it."

P11 who also continues to work in the region, described his emotional state when he went to the region for search and rescue operations on the second day after the earthquake as,

"There was chaos there, it was really bad. I can't find my tongue since the memories are bad. I was psychologically affected a lot when I saw different people and different lives coming out of the wreckage there. I felt very bad."

3.2. Process of educational coordination in the earthquake zone

In the process of educational coordination in the earthquake zone, two important points draw attention: how the situation immediately after the earthquake and the current situation are coordinated and the turnover of teachers providing services in the region. Some points such the coordination process in the region, problems and the current situation, the working environment and its effects were emphasized by the participants. For instance, P7, who went to the region in the third week after the earthquake, stated that he observed the coordination problems, *"Actually, there was a huge lack of coordination in terms of special education. The system was not fully established."* and P4 defined the coordination system established in the region where she was present with the following words,

"Each province has a coordinator. For example, there was a coordinator of PCG, there was a coordinator of special education, we also had a special education coordinator. But since he was the coordinator of Islahiye district only, he has not got individual information about all

students. He only had general information about how many students there were in the container city."

From P11's statements, who had been in the region before for the search and rescue operations and went there for the second time as a SE teacher and is currently serving in the region, it is possible to see that the working environment and the system have already been improved,

"It has started to progress more systematically compared to the first time I came here. Classes have been started in tent cities and container cities. It is systematic now. There were no materials the first time I came. Now, there are a lot of materials. It's much better in every respect."

Teacher turnover is another situation related to educational coordination, which is shaped around the temporary teachers in the region and how children with special needs are affected by this. P10 stated that teachers who came voluntarily left the region shortly after arriving and it is continued in this way, *"There was teacher turnover. A teacher comes here, stays for a week or so, then leaves."* P4 mentioned the effect of this turnover on children with special needs as,

"Actually, they were children who could progress very quickly. But there was no permanent teacher. They were very affected by the change of teachers. Children get used to seeing strangers coming and going all the time. There was someone else a week ago, and there will be another one the next week."

Later in the interview, the same participant stressed the challenging effect of teacher turnover on children with special needs in the following way, *"Children had behavioral problems because of the constant change of teachers, because of the lack of discipline."* Fortunately, in the interview conducted with P12 who is currently working in the region, it is seen that as time goes by the process improves in terms of decreased teacher turnover and permanent teachers starting to work in the region.

"The number of volunteer teachers has decreased, and the Ministry of National Education has started to hire paid teachers. Fixed-wage teachers now work in most places. So turnover has decreased."

3.3. Educational activities in the shadow of the earthquake

Among the efforts to plan and deliver education under the influence of the earthquake and the events experienced, the participants carried out a kind of screening and evaluation process through their own efforts to identify children with special needs. P1 expressed this situation as follows,

"We determined students by our efforts. We were scanning the tents one by one, together with a psychological counselor. And if there was an individual with special needs in one tent, we were taking his/her report information, which tent was he/she in, family information, phone number, and his/her name. We did such scans to find out how many students were in a tent city, how many students should receive services, how many of them had a serious condition, etc."

It is clearly seen from P4's statements that the participants wanted to start education by obtaining as much information as possible from the families of the children they reached,

"When families brought their children for the first time, they used to tell us, their child knew this and that, could not do this and that, it would be good for them if we worked in this direction, etc. They used to express their wishes in this way. So we made evaluations and started accordingly."

Furthermore, it is understood from P5's statements that there were also SE teachers who were trying to deliver educational services to the tents where children with special needs were staying since these

children had difficulties accessing educational services and their families could not bring them to the points of SE services. He said,

> "We had one student in the early childhood period. He was in a wheelchair, and his mother could not bring him because he had severe epileptic seizures. Actually, his mother could not leave her other child in the tent alone and come. Hence we started something called education services in the tent. Our special education teachers working on a rotating basis, went to the families' tents to serve. While one teacher was working in a special education tent, the other teacher was delivering service to the children, whom we would provide two hours of education during the day, in their own tents."

It is also seen that the SE teachers who participated in the study worked on what the situation there prioritized when working with children with special needs. In this sense, the participants also provided psychosocial support activities. For example, P6 stated that she also worked on how the families should behave to their children with special needs due to the adverse effects of the earthquake which made education and other life situations more difficult, in the following way, *"We took families from time to time and conducted parent trainings on how they could behave in this process."* This was sometimes requested by families, like "*There were parents who wanted psychological support*" (P6) and sometimes, it was performed to respond to a need arose spontaneously of some of the children, *"We wanted to reassure them. 'Come, my child. It is quite safe here, a truck passed outside, this is the sound of it. This place is solid. It won't be destroyed.' We tried to provide that reassurance"* (P3)

Social and psychological factors that negatively affect education include negativities that impact on the performance of participants in the region. These can be listed as the social and economic deprivation of the people in the region, the ongoing fear of earthquakes, the decrease in teacher motivation when the neat organization cannot be provided, and the negative impact of the teacher turnover on children with special needs. P8 expressed the problem that prevented a young child with special needs from participating in education as,

> "He was 3 years old and had hearing impairment. We worked with him there together with teacher T. We had taken that child once or twice before. Later, when we called him with an appointment, his mother said she was sick and could not bring him. So we said that we could take him if she wanted. My colleague T. went to take the child. When he went there, the mother said that she hadn't sent him because he didn't have socks and a coat and she was ashamed to say it."

P1 on the other hand emphasized that children did not come to receive education due to the fear of earthquakes with the following words, *"Many children were not coming due to the effects of earthquake. We worked at school, which was a building not a tent. So families were afraid to send their children, and children also were afraid to come."* P9's statement regarding the negative impact of the lack of order in the tent city on educational activities and teachers' motivation is noteworthy,

> "The fact that the tent areas were not established in a certain system was very demoralizing for us. So, when there was a fight or any other incidents in the camp, in the tent area, etc., we were very affected by it."

P7 also stated that children with special needs could not progress at the desired level due to the short stay of volunteer teachers coming to the region and frequent teacher changes, *"The fact that teachers change every week is very bad for students because students have a constant fear of separation. In this case, the provided training is not effective."*

Settling of the neat system in the earthquake zone after a while is also among the very important findings. In the interview we conducted approximately five months after the earthquake, P12 stated that the problematic situations in the system in the region decreased and this had positive reflections as

follows, *"A system has been already established. We take children on an individual basis. So their families are also satisfied."* P6 also indicated that as SE teachers continued their work, they received positive feedback from families and they saw the reward of their work in children with special needs as *"The families were praying for and thanking to us all the time."*

3.4. Children of the earthquake

Under this superordinate theme, almost all of the participants talked about the different ages, types and severity of disabilities of the children with special needs that they worked. They also emphasized their experiences with children with TD. P6 expressed the types of disabilities of the children she worked with as follows, *"There were children with autism and Down syndrome, children groups with intellectual disabilities, children with attention deficits and hyperactivity disorders, two students with hearing impairment."* while P5 indicated the severity of the disabilities of the children in the following way, *"There were children with severe, profound disabilities. We had one young student like this."* P3 stated that there were refugee children with special needs as well as Turkish citizen children by saying,

"There were five to ten children with special needs among refugees." The emphasis of P10, who also worked with children with TD in the region due to the lack of activities for these children for sometimes is noteworthy,

> *"Nothing was available at that time for normal children, typically developing children, children of preschool age. Then we tried to do something about it too. At least until something was established for them, we also tried to serve these children."*

The participants of this study also revealed the ways of children with special needs in terms of how their expressing the earthquake. P3 talked about an incident he experienced while working with children with special needs as follows,

> *"While we were inside, let's say a garbage truck was passing by outside, children were hiding under the table. It was the effect of the earthquake. Another thing with the legos. We were playing with the legos and children were shaking them and saying, 'There is an earthquake, there is an earthquake."*

The following words of P5 concerning the situation of children with special needs, particularly the little ones, developing challenging behaviors after the earthquake and the reflections of their negative experiences on their relationships with their parents are also noteworthy:

> *"Children in early childhood generally have a tremendous dependency on their parents due to the earthquake. Situations such as children crying when they go to school or crying when they leave home are observed even in normal children due to their age. But it was something we encountered very frequently in special education. After the earthquake, this situation in such children was generalized to the point of developing reactions and displaying challenging behaviors when their parents came out of the tent and reached a state that could not slow down. In other words, the mother goes out of the tent to get food, the child knocks the tent down, wondering where his mother is going. Or he would go together with his mother. These kind of behaviors are strongly linked to the fear of staying away from parents."*

Another participant, P7 also shared the challenging behaviors he observed such,

> *"I had an 8-years-old student. He had mild autism. Before the earthquakes, this student used to exhibit behaviors close to normal according to his mother. But afterwards, he suddenly withdrew into himself, he has turned into a child who has withdrawn even from his parents and does not communicate with them. He had only one toy. He has turned into a child who continuously plays with it, does nothing else. He didn't want to go into the tent, he didn't want to go to the playground, he didn't want to get involved in anything."*

4. Conclusion and Discussion

In the present study, we examined the experiences of SE teachers who went to the region to provide services in the provinces affected by the Kahramanmaraş-centered earthquakes that occurred on February 6, 2023, and worked or are currently working especially with children with special needs in the region. The phenomenon that we focused on specifically was SE teachers' experiences of working with young children with special needs. However, the interviews' results showed that the participants provide services not only to all age groups and children with special needs but also to children with TD. Therefore, we also discussed the results of this research upon this situation.

It is seen that the SE teachers participating in the study, experienced some important difficulties in the regions where the earthquakes that affected 11 provinces, but their presence in the region led them to gain some personal and professional acquisitions. Chen et al. (2022) stated that teachers working in SE classes should increase their knowledge of the dangers of earthquakes and the needs and requirements of students. The SE teachers who participated in our study noted that they did not know what they have faced when they first came to the region and were worried about this. The statement of P6, "*Due to the situation I encountered when I arrived there, I saw what a limited being I was and how limited the areas of my profession actually were.*" shows how tangible the situation in the region is and to what extent teachers actually need support in improving their support and intervention skills after the earthquake for children with special needs. As it is clearly pointed out Alisic et al. (2012) study, teachers might experience difficulties and uncertainty in their roles and responsibilities regarding how to help students suffered from traumatic stress. On the other hand, in our study, SE teachers in the region indicated the positive contribution of their experiences in working with children with special needs. Teachers clearly stated that it provided them with professional and personal skills and acquisitions. Nurses, the participants of another study showing that individuals who worked in the region developed some professional and personal positive awareness after a disaster with extremely dramatic consequences like the earthquake we experienced, stressed that their experiences inevitably changed their lives (Sloand et al., 2012).

After the Marmara earthquake in 1999 and the Van earthquake in 2011 in our country, the professionals from all disciplines were sent to the disaster regions in the quickest way and thus various psychosocial services were provided. After these disasters, it was revealed that progress was made in interventions for children in the risk group and individuals with special needs, but it was still poor, not at the desired level, and general coordination problems were experienced (Altun, 2016). Bulk of studies reported that individuals with special needs are often forgotten and neglected after natural disasters and are more negatively affected by such situations than other individuals. These studies revealed the lack of an appropriate post-disaster planning (Bilik & Akdağ, 2023; Hemingway & Priestley 2006; Masten & Osofsky, 2010; White, 2006). A recent study on the Kahramanmaraş earthquakes emphasized the need for cooperation and coordination with individuals and institutions in order to maintain the continuity of education after the earthquakes (Arıcı et al., 2023). In our study, SE teachers indicated that there were coordination deficiencies and problems when they first went to the region and SE services was not yet established in a system in some places. The participants who worked voluntarily in the region highlighted the fact that especially in the first days, there was chaos due to the coordination problems in the services provided and they had to work in that chaos. In addition to the difficulties of working in the earthquake zone, it is clearly seen that the problems experienced in planning and maintaining services reduced SE teachers' motivation to work. As Polat and Sarıçam (2024) also found, teachers face challenges that affect motivation and job satisfaction. More neat and practical system was established in the region as time passed though, educational activities were also positively affected by this, and both the places where educational activities were carried out and the duties of teachers providing services became consistent. This improvement over time is clearly observed under the superordinate theme of

educational activities in the shadow of the earthquake. In other words, as the coordination process becomes systematic and predictable, its reflection on the field is positive.

The constant change and turnover of teachers who went to the region to take part in the education process of children with special needs is another finding obtained from the study. The effect of this turnover on both teachers and children with special needs is mentioned by the participants. Other studies also report situations such as the inability to enter a regular education process due to teacher turnover and children's being unwilling to be separated from teachers after they have got used to them. In the study by Ronoh et al. (2015), children with special needs as participants pointed out that they felt safe when they are with their SE teachers. Teacher turnover adversely effects both the students and the parents (Gibbons et al., 2021). So, it might said that it also has a significant impact on an increase in challenging behaviors of children with special needs and their long-term adaptation problems after the earthquake. In addition to the personal and medical information of children with special needs, their individual education plans may also be lost after the disasters (Stough et al., 2020). The loss of these records made it difficult for teachers to reach children with special needs in an ongoing turnover and create beneficial and suitable educational plans. In the current situation, the two SE teachers who are still working in the region stated that teacher turnover decreased and permanent teachers are now working in the region. Children with special needs continue their education within a predictable environment, the system in terms of the educational services has started to be established, and families are satisfied with this. The decrease in teachers' turnover is positively reflected on children with special needs and their families.

The study revealed that the participants made efforts to identify the children with special needs in the earthquake zone and that they evaluated their educational performances after identifying the children. This situation postponed the time for children with special needs to start education. Such children became even more disadvantaged since these stages, which take time even in the normal flow of life, had to be reconstructed due to the earthquake. It is seen that the absence of planning about what should be done after natural disasters for children with special needs (Masten & Osofsky, 2010) made it difficult for them to benefit from educational services. The fact that SE teachers started providing education services at the families' tents for children who had problems reaching out the education tents caused them to make more effort and increased their workload. One mother's (P8) expression about she could not send her child to education because her son did not have socks and a coat, and she was ashamed to say it, indicated that families could not express this kind of things. The fact that families affected by the earthquakes did not send their children to the education tents because they were ashamed to clearly say that their children did not have clothes is similar to the events after the 2011 Van earthquake (Bilik & Akdağ, 2023). Although 12 years have passed, this situation has repeated. Of course such an unexpected and unfortunate event makes everything very hard and being organized for supplying food, clothes and accommodation is almost improbable. Nevertheless, to prevent such a situation, one should know and be aware how important planning in advance for what to do after natural disasters is.

It was also found that SE teachers, especially the first volunteer groups who went right after the earthquakes, carried out psychosocial support also. Teachers who observed that children also needed psychological support before their academic education needs shaped their work in this direction. This shows that children with special needs and their families have the priority in terms of a need of psychological support. We can say that the difficulty experienced by SE teachers who had to go beyond the definition of their profession originated from a lack of planning for children with special needs in natural disasters (Masten & Osofsky, 2010).

The quick return of young children to school and social settings after a disaster and return to their normal lives ensure a harmonious relationship between children's self-regulation skills and social interaction skills (Masten, 2007). In situations outside the normal flow of life, schooling is one of the

most important events for the recovery of society, returning to its normal life, and the construction of new life (Ager et al., 2010). The participants of our study worked with children with TD as well in the region. Our findings clearly shows that the SE teachers had to provide services to these children besides the children with special needs. In this way, they supported all children in the region to receive the necessary services before the school starts.

The unwillingness of young children to speak about the traumatic events they have experienced can be interpreted as that children are not affected or that they cope with the trauma more easily by society since they are too young and unaware of the events. The findings of our study are also supported by other studies reporting findings similar to the situation mentioned (Gökler Danışman & Okay, 2017; Schonfeld 2002). In other words, although children do not speak about earthquake, they are aware of and are affected by it. Fisher (2000) stated that children could recreate events that might have traumatic effects through play or drawing. In line with teachers' views, we found that children with special needs, as well as children with TD, recreated the earthquake while playing games and included it in their conversations. All children including the ones who have special needs experiencing the effects of traumatic events such as earthquakes through play. The event of recreating the earthquake by shaking legos, clearly expressed by P3, is a striking example of this.

All participants stated that some changes were observed in social skills, communication and daily living skills of children with special needs and that challenging behaviors emerged and/or increased after the earthquakes. These findings of the study coincide with the findings of research mentioning the effect of some traumatic experiences after such natural disasters (Beauchesne et al., 2002; Kurt & Gülbahçe, 2019; Ronoh et al., 2015; Valenti et al., 2012; Williams et al., 2008). According to the results, the behaviors of children with special needs differed before and after the earthquakes and challenging behaviors increased. Our findings emphasized that there has been an increase in challenging behaviors of children with special needs after the earthquake. This is an expected situation after traumatic events (Beauchesne et al., 2002; Deering, 2000; Kurt & Gülbahçe, 2019; Masten & Osofsky, 2010; Valenti et al., 2012) and it should be considered in the planning that should be prepared for the intervention for children with special needs.

5. Recommendations

Based on the findings of the study, the following recommendations have been put forward:

- Upon evaluating all the results of our study, we think that for the efficiency and effectiveness of studies after natural disasters, relevant official bodies and non-governmental organizations should create prepared teams, considering the problems caused by teacher turnover in disaster areas.
- Pre-disaster planning need to be considered among the precautions that should be taken to solve the problems of SE teachers in terms of reaching children with special needs and their records and prevent loss of time.
- Moreover, we recommend providing professional support regarding the psychological adaptation difficulties that teachers who work actively in the region may experience during the process and teachers who have returned from the region upon their return and providing disaster training to teachers and preservice teachers within the undergraduate and other education processes.
- Considering that families of children with special needs are more affected by natural disasters compared to other families, it is also important to plan and provide training to families on disaster preparedness and post-disaster processes.
- Since it is obvious that the physical and psychological difficulties of the Kahramanmaraş earthquakes and the need for support in the region will continue for a long time, we believe that

all research to be conducted in the region will guide the practical work of the relevant institutions.

- To examine the duration and status of the difficulties identified as a result of this study, it would be useful to conduct longitudinal research on the experiences of teachers from all disciplines, family members and, if possible, children with special needs themselves.

References

Afet ve Acil Durum Yönetimi Başkanlığı (AFAD) [Disaster and Emergency Management Presidency]. (2023). *Kahramanmaraş'ta Meydana Gelen Depremler Hk. [About the Earthquakes in Kahramanmaraş]*–34. Retrieved from https://www.afad.gov.tr/kahramanmarasta-meydana-gelen-depremler-hk-34

Ager, A., Stark, L., Akesson, B., & Boothby, N. (2010). Defining best practice in care and protection of children in crisis-affected settings: A Delphi study. *Child Development, 81*, 1271–1286. https://doi.org/10.1111/j.1467-8624.2010.01467.x

Aile, Çalışma ve Sosyal Hizmetler Bakanlığı. (2023). *Engelli hakları ulusal eylem planı 2023-2025 [National Action Plan for Disability Rights 2023-2025]*. Retrieved from https://www.aile.gov.tr/media/133056/engelli_haklari_ulusal_eylem_plani_23-25.pdf.

Akman, Y., & Gündoğdu, M. (2002). Türkiye-Marmara depremi sonrası gerçekleştirilen psikolojik destek çalışmaları [Psychological support studies after the Turkey-Marmara earthquake]. *Qafqaz Üniversitesi Dergisi, 1*(9), 1–7. Retrieved from https://docplayer.biz.tr/4851176-Turkiye-marmara-depremi-sonrasi-gerceklestirilen-psikolojik-destek-calismalari.html

Alexander, D., Gaillard, J. C., & Wisner, B. (2012). Disability and disaster. In B. Wisner, J. C. Gaillard, & I. Kelman (Eds.), *Handbook of hazards and disaster risk reduction,* 384–394. Routledge

Alisic, E., Bus, M., Dulack, W., Pennings, L., & Splinter, J. (2012). Teachers' experiences supporting children after traumatic exposure. *Journal of Traumatic Stress, 25*(1), 98–101. https://doi.org/10.1002/jts.20709

Altun, F. (2016). Afetlerde psikososyal hizmetler: Marmara ve Van depremleri karşılaştırmalı analizi. [Psychosocial services in disasters: A comparative analysis of Marmara and the Van Earthquake.] *İstanbul Sabahattin Zaim Üniversitesi Sosyal Bilimler Dergisi, 4*(8-9), 183–197. Retrieved from https://dergipark.org.tr/tr/pub/izusbd/issue/54076/730144

Anadolu Ajansı (AA) [Anadolu Agency]. *60 saniyede bugün (06 Şubat 2023) [Today in 60 seconds (February 06, 2023)].* Retrieved from https://www.aa.com.tr/tr/60-saniyede-bugun/60-saniyede-bugun-06-subat-2023/2808491#.

Arıcı, F., Bozkaya, H., Cengiz, E., & Kuzey, M. (2023). Education in disaster situations: The impact of the kahramanmaraş earthquake on teachers' experiences. *Sakarya University Journal of Education, 13*(4), 650-684. https://doi.org/10.19126/suje.1375691

Aslangiri, F., & O. Açkalın. (2019). Engellilerin afete dirençliliği ve Van'da yerel örgütlenme [Disaster Resilience of People with Disabilities and Local Organization in Van]. *Resilience, 3* (1), 71–83. https://doi.org/10.32569/resilience.543380

Beauchesne, M. A., Kelley, B. R., Patsdaughter, C. A., & Pickard, J. (2002). Attack on America: Children's reactions and parents' responses. *Journal of Pediatric Health Care,* 16(5), 213–21. https://doi.org/10.1067/mph.2002.127353

Bilik, M. B., & Akdağ, M. (2023). Özel gereksinimli bireyler ve ebeveynleri açısından afet zorlukları: 2011 Van depremi örneği [Disaster challenges for individuals with special needs and their parents: The case of the 2011 Van earthquake]. *Afet ve Risk Dergisi,* 6(1), 243–256. https://doi.org/10.35341/afet.1233201

Chen, Y. F., Ma, K. C., Lee, M. H., & Chuang, M. H. (2022). Earthquake Response for Students with Different Severe Degrees of Disabilities: An Investigation of the Special Education Classes in Primary

Schools in Taipei. *International Journal of Environmental Research and Public Health*, *19*(14), 8750. https://doi.org/10.3390/ijerph19148750

Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Pearson Education, Inc.

Creswell J. W., & Miller, D. L. (2000) Determining validity in qualitative inquiry. *Theory into Practice, 39*(3), 124–130. https://doi.org/10.1207/s15430421tip3903_2

International Federation of Red Cross and Red Crescent Societies [IFRC]. (2023). *Revised emergency appeal Türkiye, Europe Region |Earthquakes.* Retrieved from https://adore.ifrc.org/Download.aspx?FileId=685404

International Federation of Red Cross and Red Crescent Societies [IFRC]. (2024). *Stories of resilience in Türkiye: Rebuilding livelihoods after the earthquakes.* Retrieved from https://www.ifrc.org/article/stories-resilience-turkiye-rebuilding-livelihoods-after-earthquakes

Fisher, D. (2000). Preventing childhood trauma resulting from exposure to domestic violence. *Preventing School Failure: Alternative Education for Children and Youth, 44*(1), 25–27. https://doi.org/10.1080/10459880009602740

Foundation for the Raising and Protection of Children with Intellectual Disabilities [ZİÇEV]. (n. d). *Özel gereksinimli depremzede çocuklara destek projesi [Support Project for Earthquake-Affected Children with Special Needs].* Retrieved from https://www.zicev.org.tr/ozel-gereksinimli-depremzede-cocuklara-destek-projesi

Gazi Üniversitesi. (2023). *Bir afet bin çözüm: Özel gereksinimli öğrencileri deprem için nasıl hazırlamamız gerektiğini öğreniyoruz [One Disaster, a Thousand Solutions: Learning How to Prepare Students with Special Needs for Earthquakes].* Retrieved from https://gef.gazi.edu.tr/view/news/293098/bir-afet-bin-cozum-ozel-gereksinimli-ogrencileri-deprem-icin-nasil-hazirlamamiz-gerektigini-ogreniyoruz-isimli-tubitak-2237-a-projesi-gerceklestiriliyor

Gershon, W. S. (2009). Introduction: Working together in qualitative research-the many faces of collaboration. In W. S. Gershon (ed.). *The qualitative turn: Working together in qualitative research* (pp. xvii-xxix), Sense Publishers.

Gibbons, S., Scrutinio, V., & Telhaj, S. (2021). Teacher turnover: Effects, mechanisms and organisational responses. *Labour Economics, 73*, 102079, https://doi.org/10.1016/j.labeco.2021.102079.

Giorgi, A. (1997). The theory, practice, and evaluation of the phenomenological method as a qualitative research procedure. *Journal of Phenomenological Psychology, 28*(2), 235–260. https://doi.org/10.1163/156916297X00103

Deering, C. G. (2000). A cognitive developmental approach to understanding how children cope with disasters. *Journal of Child and Adolescent Psychiatric Nursing, 13*, 7–15. https://doi.org/10.1111/j.1744-6171.2000.tb00070.x

Gökler Danışman, I., & Okay, D. (2017). Afetlerin çocuk ve ergenler üzerindeki etkileri ve müdahale yaklaşımları [Disasters, children and adolescents: The impacts and psychological interventions]. *Türkiye Klinikleri Journal of Psychology-Special Topics, 2*(3), 189–97. Retrieved from https://www.turkiyeklinikleri.com/article/tr-afetlerin-cocuk-ve-ergenler-uzerindeki-etkileri-ve-mudahale-yaklasimlari-80107.html

Hashemi, M. (2018). Emergency evacuation of people with disabilities: A survey of drills, simulations, and accessibility. *Cogent Engineering, 5*(1), 1506304. https://doi.org/10.1080/23311916.2018.1506304.

Hemingway, L., & Priestley, M. (2006). Natural hazards, human vulnerability and disabling societies: A disaster for disabled people?. *Review of Disability Studies: An International Journal, 2*(3), 1-13. Retrieved from https://www.rdsjournal.org/index.php/journal/article/view/337/1037

Indriasari, F. N., Widyarani, L., & Daniyati, K. P. (2018). Disaster risk reduction and emergency preparedness for children with autism in facing earthquake disaster in YOGYAKARTA. *Journal Medicoeticolegal dan Manajemen Rumah Sakit, 7*(1), 52–59. https://doi.org/10.18196/jmmr.7156

İşçi, C. (2008). Deprem nedir ve nasıl korunuruz [What is an earthquake and how can we be protected from it]. *Yaşar Üniversitesi E-Dergisi, 3*(9), 959–983. Retrieved from https://dergipark.org.tr/tr/download/article-file/179169

Koç, M., & Yalçın, S. (2023). Afetlerde Krize Müdahale: Kahramanmaraş Depremi'nde Aile ve Sosyal Hizmetler Bakanlığı'nın Çalışmaları [Crisis Response in Disasters: Efforts of the Ministry of Family and Social Services in the Kahramanmaraş Earthquake]. *Uluslararası Sosyal Hizmet Araştırmaları Dergisi, 3*(2), 93-105. https://doi.org/10.57114/jswrpub.1257945

Kothari, C. R. (2004). *Research methodology: Methods and techniques.* New Age International (P) Ltd.

Kurt, E., & Gülbahçe, A. (2019). Van depremini yaşayan öğrencilerin travma sonrası stres bozukluğu düzeylerinin incelenmesi [Investigation of post-traumatic stress disorder levels of students living in Van earthquake]. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 23*(3), 957–972. https://dergipark.org.tr/tr/download/article-file/835828

Masten, A. S. (2007). Resilience in developing systems: Progress and promise as the fourth wave rises. *Development and Psychopathology, 19*, 921–930. https://doi.org/10.1017/S0954579407000442

Masten, A. S., & Osofsky, J. D. (2010). Disasters and their impact on child development: Introduction to the special section. *Child development, 81*(4), 1029–1039. https://doi.org/10.1111/j.1467-8624.2010.01452.x

Merriam, S. B. (2013). Nitel Araştırma: Desen ve uygulama için bir rehber [Qualitative Research: A Guide for Design and Implementation] (Trans. Ed. S. Turan). Nobel Yayıncılık.

Millî Eğitim Bakanlığı (MEB) [Ministry of National Education (MoNE)]. (2023a). *MEB, deprem yaralarını sarmak için tüm birimleriyle seferberlik halinde [The MoNE is mobilized with all its units to heal the wounds of the earthquake].* Retrieved from https://www.meb.gov.tr/meb-deprem-yaralarini-sarmak-icin-tum-birimleriyle-seferberlik-halinde/haber/29161/tr

Millî Eğitim Bakanlığı (MEB). (2023b). *Özel eğitime ihtiyacı olan öğrenciler için afet sonrası uyum etkinlikleri kitabı yayımlandı.* Retrieved from https://orgm.meb.gov.tr/www/ozel-egitim-ihtiyaci-olan-ogrenciler-icin-afet-sonrasi-uyum-etkinlikleri-kitabi-yayimlandi/icerik/2336

Morgan, D. L. (1996). *Focus groups as qualitative research.* Sage Publications.

Patton, M. Q. (2002). *Qualitative research & Evaluation methods.* 3rd Edition. Sage Publications.

Polat, M. S. & Sarıçam, H. (2024). 6 Şubat depreminden etkilenen bölgelerdeki öğretmenlerin karşılaştığı sorunlar: Hatay ili örneği. *Akademik Yaklaşımlar Dergisi, 15*(1), 764–789. https://doi.org/10.54688/ayd.1412861

Ronoh, S., Gaillard, J., & Marlowe, J. (2015). Children with disabilities and disaster preparedness: A case study of Christchurch. *Kōtuitui: New Zealand Journal of Social Sciences Online, 10*(2), 91–102. https://doi.org/10.1080/1177083x.2015

Schonfeld, D. J. (2002). Supporting adolescents in times of national crisis: Potential roles for adolescent health care providers. *Journal of Adolescent Health, 30*(5), 302–307. https://doi.org/10.1016/S1054-139X(02)00382-8

Sloand, E., Ho, G., Klimmek, R., Pho, A., & Kub, J. (2012). Nursing children after a disaster: A qualitative study of nurse volunteers and children after the Haiti earthquake. *Journal for Specialists in Pediatric Nursing, 17*(3), 242–253. https://doi.org/10.1111/j.1744-6155.2012.00338.x

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage Publications, Inc.

Stough, L. M., Ducy, E. M., Kang, D., & Lee, S. (2020). Disasters, schools, and children: Disability at the intersection. *International Journal of Disaster Risk Reduction, 45*, 101447. https://doi.org/10.1016/j.ijdrr.2019.101447

Thomas, D. R. (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation, 27*(2), 237–246. https://doi.org/10.1177/1098214005283748.

Türkiye Cumhuriyeti Cumhurbaşkanlığı Strateji ve Bütçe Başkanlığı (SBB) [Presidency of the Republic of Turkey, Presidency of Strategy and Budget (PoSB)]. (2023). *2023 Kahramanmaraş ve Hatay depremleri raporu [2023 Kahramanmaraş and Hatay earthquakes report]*. Retrieved from https://www.sbb.gov.tr/wp-content/uploads/2023/03/2023-Kahramanmaras-ve-Hatay-Depremleri-Raporu.pdf

Söğüt, D. A., & Kaya, A. (2024). 'In order not to cry, we averted our eyes from each other': Experiences of special education professionals working in the field during the Turkey earthquake, *International Journal of Developmental Disabilities,* (early view*),* https://doi.org/10.1080/20473 869.2024.2324529

Valenti, M., Ciprietti, T., Di Egidio, C., Gabrielli, M., Masedu, F., Tomassini, A. R., & Sorge, G. (2012). Adaptive response of children and adolescents with autism to the 2009 earthquake in L'Aquila, Italy. *Journal of Autism and Developmental Disorders, 42*(6), 954–60. https://doi.org/10.1007/s10803-011-1323-9

Van Manen, M., McClelland, J., & Plihal, J. (2007). Naming student experiences and experiencing student naming. In D. Thiessen & A. Cook-Sather (Eds.). *International Handbook of Student Experience in Elementary and Secondary School* (p. 85-98). Springer. Retrieved from https://link.springer.com/chapter/10.1007/1-4020-3367-2_3

White, B. (2006). Disaster relief for deaf persons: Lessons from hurricanes Katrina and Rita. *Review of Disability Studies: An International Journal, 2*(3). Retrieved from https://www.rdsjournal.org/index.php/journal/article/view/336/1034

Williams, R., Alexander, D. A., Bolsover, D., & Bakke, F. (2008). Children, resilience and disasters: Recent evidence that should influence a model of psychosocial care. *Current Opinion in Psychiatry, 21*, 338–344. https://doi.org/10.1097/YCO.0b013e328305b6e4

World Federation of Occupational Therapists (WFOT). (2007). *Strategic plan 2007-2012*. Retrieved from http://www.netf.no/internasjonalt/WFOT

Yarımkaya, E., & Bakkaloğlu, H. (2024). Experiences of parents of individuals with intellectual disabilities after the 2023 Kahramanmaraş earthquakes in Türkiye: A phenomenological study.

Emine Ayyıldız, Beyza Elif Çiçen, Funda Ün, Hikmet Taha Acartürk, Nisa Demir

Article Information Form

Authors Notes: This study was presented at 7.Uluslararası Katılımlı Ulusal Disiplinerarası Erken Çocuklukta Müdahale Kongresi (UDEMKO2023), Eskişehir, Turkiye.

Authors Contributions: All authors contributed equally to the writing of this paper. All authors read and approved the final manuscript.

Conflict of Interest Disclosure: No potential conflict of interest was declared by authors.

Copyright Statement: Authors own the copyright of their work published in the journal and their work is published under the CC BY-NC 4.0 license.

Supporting/Supporting Organizations: No grants were received from any public, private or non-profit organizations for this research.

Ethical Approval: It is declared that during the preparation process of this study, scientific and ethical principles were followed and all the studies benefited from are stated in the bibliography. The ethics committee document for this study was obtained from the Istanbul Medeniyet University Educational Sciences Ethics Committee with the decision numbered 2023/04-09 dated 08.05.2023.

Plagiarism Statement: This article has been scanned by iThenticate.

Use of Artificial Intelligence (AI) Technologies in Education According to Primary School Teachers: Opportunities and Challenges

Mustafa Erol[1*] (ID)
Ahmet Erol[2] (ID)

[1] Yildiz Technical University, İstanbul, Türkiye, merol@yildiz.edu.tr

[2] Pamukkale University, Denizli, Türkiye, ahmete@pau.edu.tr

*Corresponding author

Abstract: Artificial intelligence (AI), which refers to technologies that mimic human cognition, affects many industries. Education is one of these sectors. Artificial intelligence affects many educational environments, from lectures to homework. In this process, both academic and ethical concerns call into question the future of artificial intelligence. These inquiries are essential as they show that the human factor will continue as an integral part of education. Because AI tools, even when best designed, can only partially replace human interaction or quality teaching. However, they can make the teacher's job easier and contribute to more effective learning. Therefore, teachers' awareness of this technology has become essential. This research aims to determine primary school teachers' opinions about using AI tools in education. The research was conducted using a case study. The participants are 16 primary school teachers determined by the criterion sampling method. Data were collected through a semi-structured interview form and analyzed with content analysis. According to the findings, teachers stated that AI tools may have advantages and disadvantages in educational environments. While teachers are concerned about the adverse effects of artificial intelligence tools on students, they also recognize their cognitive and socio-emotional contributions. Teachers also stated that artificial intelligence can make teachers' jobs easier but can only partially replace them. The results help understand primary school teachers' opinions regarding using artificial intelligence tools in the learning process.

Keywords: Artificial Intelligence, AI Tools, Artificial Intelligence in Education, Primary School Teachers

1. Introduction

Today, digital technologies are turning a life that was once considered a dream and described in detail by science fiction writers into reality. Authors such as Asimov (1950) pioneered us to start living these dreams today by establishing these dreams that shape the present and the future. John McCarthy introduced the concept of artificial intelligence (AI) in 1956. With technology developing since 1956, AI has found a more significant place in our lives. Significantly since the early 2000s, the advancement of digital technologies, the internet connecting millions of people and providing easy access and sharing of information, has radically changed human life and pushed people to different pursuits (Erol et al., 2023; Luckin & Cukurova, 2019). Ways of accessing and using information have evolved rapidly due to the rapid advancement of digital technology and increasing knowledge. This evolution has led to AI technologies and tools becoming more prevalent. As in every field, education and training activities have been affected by AI technologies (Zang & Aslan, 2021).

In 1956, John McCarthy described artificial intelligence (AL) as "the science and engineering of creating intelligent machines" (McCarthy, 2007). Mondal (2020) defined AI as the science and engineering of producing intelligent machines that solve different problems through natural language processing, neural networks, and machine learning. In other words, AI can be defined as an information technology that can imitate human cognition. AI represents the capacity of computer tools to perform tasks that often involve human intelligence, such as learning, reasoning, problem-solving, and decision-making (Minkkinen & Mäntymäki, 2023). According to Gondal (2018), AI has a skill set that can perceive human cognition, reason, comprehend, make sense of, generalize, infer, learn, and perform multiple tasks simultaneously.

AI technologies aim to personalize learning materials based on students' abilities, preferred learning styles, and experiences. In this respect, researchers have recently been discussing the reflections of AI on education and training activities (Bayraktar et al., 2023; Eren, 2021; Zhang & Tur, 2023) and

researching the use of this technology with young children (Su et al., 2023; Yang, 2022; Su & Yang, 2023). This change points to a new paradigm in learning and teaching methods. Within the framework of this changing paradigm, this study focuses on the advantages, opportunities, challenges, and obstacles of using AI in the primary school period from the perspective of classroom teachers. For this purpose, the theoretical part first discusses the phenomenon of AI in education, discusses the advantages, expectations, difficulties, and obstacles of using AI in education in the context of the literature, examines the place of AI in teaching activities from the perspective of classroom teachers, and ends with the importance of the study.

1.1. The age of artificial intelligence in education: Shaping the future

In the digital age, AI is reshaping many industries. It is possible to express one of these sectors as education. In education fields, AI can help teachers predict students' learning status and performance, recommend learning resources, and improve students' learning experience with intelligent teaching systems (Liang et al., 2021; Zheng et al., 2021). The focus of AI in the field of education is AI literacy. AI literacy refers to competencies that enable people to critically evaluate, communicate, and collaborate with AI, as well as being an end user of AI tools (Druga & Ko, 2021; Holmes et al., 2022). However, research focuses little on improving AI literacy in children and its effects (Su et al., 2023). Children may need to learn how to use AI and the basic working principles behind these tools and may need clarification about the technologies (Su et al., 2023). Recently, the development of more age-appropriate software has allowed children to expand their opportunities to learn and explore AI and improve AI literacy (Yang, 2022). Therefore, teachers' professional development of this new technology is essential in their learning processes.

Today, young children are growing up with AI-powered technologies. AI-supported technologies have begun participating in children's lives in every field, from computer games to socialization environments. Focusing on this early interaction with AI systems, researchers have begun to examine the possible effects of AI technologies on children (Edwards, 2023; Williams et al., 2019). At this point, the literature questions whether preschool and primary school students are too young to discover and learn AI knowledge (Su et al., 2022). However, previous studies have shown that AI technologies are promising for children (e.g., Lin et al., 2020; Tseng et al., 2021; Williams et al., 2019). AI education has brought challenges and opportunities to early childhood education, including why children should learn AI in their early years, the subset of basic AI concepts that children can understand, and meaningful learning for children (Yang, 2022). In this respect, since AI is now part of the 'real' world in individuals' lives, including the early years, it is time to consider how and why information is important and how people will use these technologies in the post-digital age (Edwards, 2023).

1.2. Children in the age of artificial intelligence: Opportunities and challenges

AI offers numerous possibilities for innovation, especially in early childhood education (Lomier, 2023). With the increasing number of developmentally well-designed technologies for young children, young children can discover AI through playful experiences (Su et al., 2023). Interaction with AI can help children improve their digital literacy skills and attitudes and accelerate their progress in primary school. However, it is acknowledged that young children have little knowledge or understanding of AI (Su et al., 2023). Supporting students' cognitive development can help develop motor skills, aid speech and language acquisition, and enhance social-emotional learning, which can fundamentally change our approach to early learning (Lomier, 2023; Su & Zhong, 2022). Parents can use AI toys and services to create a digital environment for their children, and this new skill may make them more comfortable using such tools and increase the likelihood that they will use it as part of their instructional designs in the future (Knox, 2019; Su et al., 2023).

Although studies on AI for young children are still in their infancy, researchers have begun to explore how AI applications are used to facilitate student learning and the management of kindergarten and primary school teachers (PST) through intelligent tutoring systems for special education, chatbots for language education, and robotic kits in schools (Su et al. al., 2023). In early childhood education, AI-powered toys allow students to learn robots and kits. It also provides a fun experience for children to interact with them and support their coding skills. With well-designed AI systems, young children can improve their AI literacy even in the early years (Williams et al., 2019). Students with AI skill proficiency can think computationally based on their programming skills (Kim et al., 2021). Young children can explore AI technologies and improve their digital literacy in their daily lives, even if they do not know or understand the knowledge behind it. Education, health, and social opportunities for young children form part of their daily lives and function as components of broader society. AI is now part of the natural world where young children and their families live and, by extension, workplace employees (Edwards, 2023; Yi et al., 2024).

Although studies on AI focus on the many positive contributions of AI technologies for children, there are still various concerns. First, privacy and security concerns examine AI's effects on children's privacy and safety (White, 2018; Akgun & Greenhow, 2022). Identity protection, financial security, and preventing fabricated identities are critical to protecting children from AI technologies (UNICEF, 2019; Zanetti, 2020). In addition, harmful content, location detection, biological security, and the potential effects of AI on health are factors that need to be considered for the safety of children (Chou et al., 2017; UNICEF, 2021). From the perspective of access to services and equity, the risk of AI and machine learning leading to discriminatory outcomes must be considered (Ayala-Pazmiño, 2023; Kavale & Forness, 2019; Mengesha, 2021). The issues of who can access which services, how these decisions are made, and whether they are treated somewhat are essential for children's access to equal opportunities (UNICEF, 2021).

At another point, the contributions AI can make to the social development of children should be examined (Kurian, 2023; Mengesha, 2021). At the same time, potential impacts on how children spend their leisure time and how AI can promote purposeful leisure should also be addressed (White, 2018; Akgun & Greenhow, 2022). Regarding cognitive and psychological effects, the effects of AI on the brain should be examined. Issues such as the effects of this technology on children's psychological health, depression, anxiety, and the development of social skills are essential factors to consider (UNICEF, 2021; Zanetti et al., 2020). The potential risks of cognitive manipulation and the use of AI to direct child behavior should be evaluated (Chou et al., 2017).

1.3. Primary school teachers in the age of artificial intelligence: Responsibilities and expectations

According to the goals of the Ministry of National Education, AI technologies are planned to be used to improve education in Türkiye (Celik et al., 2022; Sevil & Saralar-Aras, 2024). In this context, research containing detailed information about AI's applications in education and its advantages is essential for educators to follow current developments and adapt to technology. Additionally, this research serves as a resource for anyone who aims to understand PST's views on AI-based education by providing up-to-date information (Gordon, 2011). As discussed above, AI has advantages and challenges in education and training activities. Although there are difficulties and obstacles, AI will inevitably occur among today's teaching technologies, considering its advantages. In this regard, it is an important context to examine the technology in question from the perspective of teachers in primary school classes, especially from the early years.

AI technologies, even the best-designed ones, can never replace human interaction or good teaching for the primary school age (Bulut et al., 2024; Kolchenko, 2018). However, they can serve more effective

learning by making the teacher's job easier. Children need caring and knowledgeable adults to help them navigate and learn about the world; This includes the world of technology. In this context, integrating AI technologies into educational environments without neglecting adult support will increase the quality of education (Kolchenko, 2018; Zang & Aslan, 2021). AI technologies can help teachers with many tasks, such as creating lesson plans, helping with homework evaluation, offering personalized support to students, and even creating videos from textbooks. The potential of these applications is quite wide-ranging, and PST should aim to take full advantage of these technological advances (Bowman, 2023).

This study is crucial for several reasons. Firstly, understanding PST's perspectives on integrating AI in education provides invaluable insights into the practical challenges and opportunities of implementing AI technologies in classrooms. Teachers are on the frontline of educational practice, and their experiences and opinions are vital for shaping effective AI-based educational strategies. While previous studies, such as those by Bayraktar et al. (2023), have explored AI's theoretical and broad implications in education, a distinct lack of research focuses on teachers' practical, day-to-day experiences. Secondly, this study addresses a significant gap in the current literature by providing detailed insights into the specific benefits and obstacles of AI from the perspective of PST. This focus is essential because teachers' acceptance and effective use of AI technologies are critical for successfully integrating these tools into education. By identifying teachers' specific needs, concerns, and suggestions, this research can inform the development of more targeted professional development programs and support systems to help teachers adapt to and embrace AI technologies. Lastly, the study's findings can serve as a valuable resource for policymakers, curriculum developers, and educational technology designers. By highlighting AI's real-world challenges and benefits in education, this research can guide the development of policies and tools that support the effective and equitable use of AI in primary education. This is particularly important in ensuring all students have access to high-quality AI education and are prepared to navigate the increasingly digital world.

It can be stated that teachers' perceptions of their skills in digital technologies and AI can affect the activities planned for children and AI applications (Kolchenko, 2024; Seyrek et al., 2024). In this respect, teachers must have good preparation in content and pedagogy so that children can access quality AI experiences. However, current pre-service and in-service training programs must pay more attention to AI training for PST. Determining teachers' perspectives will create the content of professional development programs for teachers and facilitate the integration of AI into primary school education environments. Therefore, this study aims to reveal PST's views on what opportunities and challenges AI technologies will create in educational environments. The questions guiding the research are presented below.

- According to PST, what are the benefits of using AI technologies for children in primary school?

- According to PST, what are the challenges to using AI technologies in primary school?

- According to PST, what is the role of AI technologies in the educational environment in primary school?

2. Method

2.1. Research model

This study used a case study to determine PST's thoughts on using AI tools. A case study is a research method used to examine a specific situation, event, group, or community in detail (Merriam, 2015). The situation examined in this study is teacher opinions regarding AI tools. Researchers often collect in-depth data during the research process to understand events or processes occurring at a particular time and place. Researchers widely use case studies to understand complex and real-life situations,

understand a specific context, and investigate cause-effect relationships (Patton, 2015). This study conducted in-depth interviews to understand better teachers' mental structures, experiences, and interactions with AI technologies (Merriam, 2015).

2.2. Participants

The study group of the research was determined by the criterion sampling method, one of the purposeful sampling types. Criterion sampling is a method in which researchers select participants based on specific criteria when determining the study group. Researchers use this method to include individuals who meet predetermined criteria in the sample. These criteria depend on the research aims, scope, and hypotheses. This method allows researchers to narrow the study group, conduct more in-depth analyses in a specific context, and reach more specific results (Patton, 2015). Determined criteria;

a) *Being a primary school teacher*: This criterion constitutes the study's main context because the focus is to examine the opinions of PST.

b) *Working in public primary schools in Istanbul*: This criterion is essential for access to the study group. Reaching teachers in other cities may cause a loss of time and effort, so the criterion for working in public primary schools in Istanbul has been determined.

c) *Knowledge about AI systems*: This criterion is essential for teachers to answer researchers' questions more effectively. Getting in-depth information from a teacher who does not know the subject is impossible. Teachers were asked the following questions: Have you attended a training on AI before? If you attended, could you give us brief information about it?

d) *Using an application that includes AI systems*: This criterion was determined to ensure teachers have gained experience with AI systems. In this way, teachers can provide more qualified and in-depth information about AI systems.

The study group consists of 16 PST's who teach in public primary schools in different districts of Istanbul. The table below presents demographic information about the study participants.

Table 1

Demographic Information of the Study Group

Variables		f
Gender	Female	9
	Male	7
Age	25-30	5
	31-35	4
	36-40	4
	41-45	1
	46 +	2
Professional Seniority	1-5	3
	6-10	5
	11-15	5
	16 +	3

2.3. Data collection

A semi-structured interview form was used since this study wanted to examine teachers' opinions about AI systems. The interview form is structured in two stages. The first stage consists of questions about gender, age, and working hours to determine teachers' demographic characteristics. The second stage

consists of a form with five questions, including semi-structured interview questions. While preparing this form, the relevant literature was examined, and expert opinion was sought. After the experts gave their views, some questions were renewed, and the final form was provided. In this context, the interview method was preferred to reveal PST's ideas and feelings about AI and whether they would replace the teacher. Sample questions in the interview form are as follows: 1) What are the advantages of using AI tools in educational environments? 2) Could AI replace the teacher one day? To increase the richness of the data in the interview questions and deepen the participants' opinions, some additional questions, described as probes, were asked by the flow of the research. Probes provide accommodations to ask more questions about details, ask for clarification, or obtain examples (Merriam, 2015). The interviews were held in a quiet environment at the schools where the participants worked and were recorded with the help of a voice recorder.

During the data collection process, PST were determined first. Necessary permissions were obtained from the designated teachers' institutions. In addition, written consent was obtained from the study participants. Among the teachers reached, 16 teachers who used AI systems in their classrooms at least once constituted the study's participants. Participants were informed about the purpose and outcomes of the research. As a result of the information provided, three teachers wanted to refrain from participating in the research. After the participants were determined, interviews were conducted using the interview form in a quiet environment (teachers' room, etc.). The interviews lasted an average of 29.8 minutes with each teacher. The phone's voice recording feature prevented data loss during the interviews. The data obtained were transcribed, and all audio recordings were examined.

2.4. Data analysis

The analysis of the data obtained through semi-structured interviews was carried out using the content analysis method. Content analysis is expressed as a dimension, such as scanning the qualitative text in line with recurring words and themes, reducing and making sense of qualitative data by taking the voluminous qualitative material, and determining its essential consistency and meaning (Patton, 2015). Coding was first done in line with content analysis, and themes were reached through this coding (Merriam, 2015). In this regard, coding was created as small notes in the study. By combining repeated coding, themes (categories) emerged (Yıldırım & Şimşek, 2013). Internal validity was ensured by examining the compatibility of the results obtained from the analysis with the previously established conceptual framework and theories. In the research, direct quotes were made from the interviews for internal reliability. After the data obtained from the interview was transcribed, it was sent to the participants in writing, and external reliability was ensured by asking whether the data accurately reflected their perceptions. In data analysis, abstraction was made using an inductive method, coding the data and reaching themes from the coding. First, all interview documents were transcribed. Transcripts were read, and each statement related to the experience was identified. All variables related to AI systems are listed. Then, variables with the same meaning were grouped, and meaning units were created. The themes were re-read after a specific interval, and new themes were obtained. In addition, the themes obtained from the analysis processes were presented under headings in the findings section and discussed with the findings of other studies in the literature.

2.5. Credibility and transferability

Two researchers conducted the coding processes in the data analysis at different times to ensure inter-coder reliability and expert opinions supported the determined themes and codes. The inter-coder agreement coefficient was examined with Kappa and was found to be .93. Care was taken to enrich the content while creating codes and themes using direct quotes from teacher interviews. Additionally, leaving 35 days between the data collection and analysis process minimized the possibility of researchers' individual opinions interfering with the analysis process. To ensure the external validity of the research, the criterion sampling method was chosen using the purposeful sampling method of the

study group. The interviews were conducted face-to-face with the teachers. Additionally, to increase the research's external reliability, a literature review was conducted to prepare the interview questions, pilot interviews were held with teachers, and the questions were reshaped based on expert opinions.

2.6. Role of Researchers and Ethics

Before starting the research, approval was obtained from the university's social and human studies ethics committee (Meeting Date: 04.03.2024; Meeting Number: 2024.03). The ethics committee document for this study was obtained from Yıldız Technical University Social and Human Sciences Research Ethics Committee with the decision dated 04.03.2024 and numbered 2024.03. The research's purpose, process, and potential risks were clearly explained to the participants, and their consent was obtained. Consent forms were filled out and signed by the participants in understandable language. It was notified in advance that the interviews would be audio recorded. The personal information of the participants was kept confidential and was used only during the research process. Participants' identities were not used in research reports or presentations. Participants' names or other personal identifying information were not shared anywhere without their consent. Researchers informed participants that they could leave the study at any time and not talk about questions they did not want to answer. Care was taken during the data collection process to conduct the interviews as objectively as possible. In this regard, care was taken not to judge the behavior and words of the participants and not to create any expectations for them. Additionally, to ensure participant control over the findings, observations were made throughout the research process, the researcher was enabled to state his prejudices, and time was spent with the participants in the same environment for a certain period.

3. Findings

This section focuses on classroom teachers' views on using AI tools in education. The data analysis findings obtained from the teachers were explained with three themes: 1) Negative Effects of AI Technologies on Students, 2) Advantages of AI technologies, and 3) The Role of AI in the Educational Environment. Below, the themes, sub-themes, and codes obtained from the teachers' opinions are presented verbatim from the teachers' statements.

3.1. Theme 1: Negative effects of AI technologies on students

This theme considers the potential adverse effects of AI tools on students through teachers' eyes. Based on the data obtained, we can say that, according to teachers, the tools may negatively affect students' health and learning processes. Teachers have stated that the widespread use of AI tools will negatively affect students' cognitive, emotional, and value development. They are especially concerned about its unconscious use. P1 emphasized this situation: "*I think that when used unconsciously, it can prevent the development of thinking ability, questioning and make people lazy.*" Teachers also think that the awareness of responsibility in students will decrease: P14 "*Failure to understand the importance of working hard for success and weakening of the awareness of responsibility.*" According to teachers, the adverse effects of AI tools on students are summarized in Table 2 through themes, sub-themes, and codes.

Table 2

Adverse Effects of AI on Students

Themes	Sub-themes	Codes	f
Adverse Effects of AI on Students	Affects Thinking Skills	Reduces creativity	9
		Socialization problems	8
		Reduces students' curiosity	8
		Inhibits reasoning	7
		It inhibits the thinking process	7
		Negatively affects your critical thinking skills	6
		Writing skills are negatively affected	6
		Idea generation decreases	3
	Socialization Problems	It causes laziness	13
		Negatively affects friendship skills	10
		The bond between child and teacher is not strengthened	10
		Social rules are not learned	9
		Students fail to realize their responsibilities	9
		The child becomes numb	7
		Negatively affects empathy development	6
		Negatively affects emotion regulation skills	5
		Negatively affects self-efficacy	4
	Worthlessness	Education of national values becomes difficult	7
		Acquisition of social values becomes difficult	6
		Values such as loyalty, friendship, and compassion are difficult to learn.	2
		A worthless society is born	1
		The importance of school disappears	8
		The importance of education decreases	7

Table 2 shows that teachers have opposing views on the use of AI tools in education. Most teachers think AI tools will reduce creativity in students, negatively affect friendship skills, reduce national values, and cause laziness. Teachers' opinions are explained with three sub-themes: 1) impacts on thinking skills, 2) socialization problems, and 3) worthlessness.

3.1.1. Affects thinking skills

Teachers think that with the increasing use of AI tools, students' creativity will decrease, social problems will become more challenging to solve, reasoning will be hindered, critical thinking skills will be negatively affected, writing skills will be negatively affected, and idea generation will decrease. According to teachers, if AI is used unconsciously, students' thinking and questioning skills may be negatively affected. P11 emphasized this situation: "*It may negatively affect students' thinking skills.*" In addition, teachers stated that humans might evolve into machines when they have an insufficient understanding of emotions, are incompatible with dynamic changes in human interaction, and ignore sincerity and emotion. A teacher expressed his views as follows.

> *P3: It may be insufficient to understand emotions during the education and training. Human interacting beings change and develop thanks to this interaction. In addition, the human mind is dynamic, and sensitive thoughts constantly change. He cannot understand this and look at the student with the same eyes and understand the student's mind. It is insufficient to react. Classical coding or mathematical approaches. This ignores sincerity and emotion. "This situation may lead humans to evolve towards mechanization.*

On the other hand, according to teachers, the concern that AI will negatively affect students' thinking skills and the thought of hurting the education process and the whole of life come to the fore. They stated that AI may cause students to succeed in homework assignments without putting in thought and effort

and that it may prevent process-oriented education by reducing research and questioning skills, which will negatively impact cognitive development. For example;

> *P1: It will hurt the whole of life along with education. "It causes people to take it easy in fulfilling their duties and responsibilities and to fulfill them in a way that is far from reaching the goal, without giving due importance, especially the inability to express feelings and thoughts accurately and impressively.*

On the other hand, teachers have also expressed concerns that AI will lead people to ready-made habits, cause the imagination to dull, the learning process may cease to be a pleasant experience, and may reduce kinesthetic skills. P3 explains this situation as follows.

> *P3: AI can get people used to the present and dull their imagination. One of the most essential characteristics of a human being is the ability to dream; if everything is done with Ali, the human development process will be blocked. The learning process is pleasurable if you do and experience it. As a person does things on his own and achieves things, his desire increases, his self-confidence rises, and his kinesthetic skills develop. These developing processes have a positive effect on other areas of life as well.*

Finally, teachers also expressed their concerns that with the widespread use of AI tools, children's analysis and synthesis skills will be negatively affected, their awareness of effort and responsibility will weaken, students will have difficulty expressing themselves, their writing skills will decline, and their thinking abilities will disappear. These statements reflect concerns that AI may have a wide range of effects on education and that these effects may have negative consequences on student development. Some teacher opinions that support these statements are as follows;

> *P2: The purpose of these assignments is to encourage students to think. The student does not do the homework using AI. It gives the impression that someone else has done his work instead of him. In this way, the desired outcome does not reach the student. Process-oriented education, which requires students to research and question, is disrupted. The contribution of the assigned work to student success is significantly reduced.*

> *P5: The participant's subjective judgments will decrease. It will also negatively affect children's analysis and synthesis skills in the long term.*

> *P16: Students may have difficulty expressing themselves. Writing skills may decline. "Thinking, ideation, and problem-solving skills may be lost.*

3.1.2. Socialization Problems

According to teachers, AI tools also bring about socialization problems for students. For example, P1's statement, "*I think that AI tools will become widespread in the near future and students will move away from society*," supports this situation. In particular, the unconscious use of AI tools brings many problems, such as negatively affecting friendship skills, difficulty learning social rules, not being aware of responsibilities, desensitization, negatively affecting empathy development and emotion regulation skills, and negatively affecting self-efficacy. P7 expressed this situation as follows:

> *P7: The spread of AI means a decrease in social relations. It adds to the countless children who are stuck at home with technology today. Do we really want this as educators?*

> *P8: With AI, children will encounter many negativities such as difficulty learning social rules, not being aware of responsibilities, desensitization, and adverse effects on empathy development. This situation will deepen with applications such as Metaverse.*

> *P2: With AI, children's emotion regulation skills will be negatively affected, and their self-efficacy will be negatively affected. Because it is a virtual world for children, they will be emotionally depressed by moving away from reality.*

3.1.3. Worthlessness

According to teachers, AI tools may make the education of national values difficult for students, the acquisition of social values may be complex, values such as loyalty, friendship, and compassion may be difficult to learn, a worthless society will arise, and the value of education and school may decrease. Values are learned formally or informally in the society we live in. AI tools will negatively affect this situation, making it difficult to transfer our cultural values to future generations. The teacher's opinion supporting these statements is as follows;

> *P9: With the widespread use of AI tools, education of national values will become difficult, acquisition of social values will become difficult, and values such as loyalty, friendship, and compassion will lose their importance because children will get used to loneliness in this process.*

> *P5: The widespread use of AI could shift societal dynamics, potentially resulting in a devaluation of education and schools.*

> *P14: "Values are essential and unwritten rules for the continuation of society. Just as their tools will reduce the importance of culture, they will also reduce the acquisition of these values. What will people want from now on in a worthless society?*

When the adverse effects of artificial intelligence are examined according to teachers' views, the following conclusions can be made. It can lead to independent thinking in education, socialization problems, difficulty acquiring national and cultural values, and weakening problem-solving skills. Although these technologies can provide students instant information and automatic responses, their constant dependence on such tools can dull their critical thinking abilities and prevent creative thinking processes. In addition, the promise of personalized learning by artificial intelligence can negatively affect the development of group work and communication skills by reducing competition and social interaction among students. In the long run, this situation can cause students to become dependent on technologies and individuals with difficulty developing a critical perspective on the information these technologies provide.

3.2. Theme 2: Advantages of AI technologies

According to teachers, this theme examines the advantages AI tools can bring to students. Teachers have stated that AI tools can enhance or support students' learning experiences in certain areas. P7 emphasized this situation: "AI is inevitable; it will offer many new opportunities. If our students can use it well, their learning experiences will expand". Another teacher stated that "AI is not actually scary, it should be seen as an opportunity to expand learning environments" (P9). The advantages of AI tools compared to teachers are summarized in Table 3 through themes, subthemes, and codes.

Table 3

Advantages of AI technologies

Themes	Sub-themes	Codes	f
Advantages of AI Tools	Cognitive Contribution	Supports learning	10
		Makes teaching easier	9
		Improve academic success	9
		Contributes to problem-solving	9
		Makes teaching reading and writing easier	8
	Socio-Emotional Contribution	Increases motivation	7
		Increases interest in the lesson	7
		Develop a positive attitude toward the lesson	7
		Reduces prejudice	6
		Used in the communication process	6
		Increases curiosity	3
	Other contributions	Provides fun time	8
		It can be used as educational material	7
		Helps students while doing homework	6
		Provides equal opportunity in crowded classes	6
		Saves time and space	5
		Speeds up access to information	4

When Table 3 was examined, teachers stated that AI positively contributes to cognitive, socio-emotional, and other fields. According to teachers, the positive effects of AI tools manifest themselves in various dimensions. These effects are explained with three sub-themes: 1) Cognitive Contribution, 2) Socio-Emotional Contribution, and 3) Other Contributions.

3.2.1. Cognitive contribution

Teachers emphasized that access to information will be easier with AI tools and, as a result, will contribute to children's cognitive development. According to teachers, AI tools can primarily support learning in the educational process, facilitate teaching, and increase academic success in children. Additionally, it contributes to students' problem-solving skills. On the other hand, according to teachers, AI can make teaching reading and writing easier. The following statements support these explanations.

P2: Al can be used for appropriate purposes. The opportunities provided by AI can be used in science, in fields that require three-dimensional work, or in situations where there is a material shortage.

P3: It should be used as an auxiliary tool at school and home. It should be used as a tool to accelerate or facilitate the student's learning process, sometimes while doing homework at home, and sometimes in the classroom, as a different method and technique.

P4: Al can be used. It provides great convenience in terms of speed of access to information. It can be used by teachers in the course planning process, course preparation process, searching for articles by subject, and providing quick access to related articles.

P16: Access to information is accelerated, education is easily supported with visual and audio materials, time and space are saved, the educational environment becomes fun, and the memorability of the acquired knowledge increases.

3.2.2. Socio-emotional contribution

According to teachers, interacting with AI tools contributes to children's cognitive development and provides many social-emotional contributions. First, AI tools can positively affect children's learning by increasing their motivation, attracting interest in the lesson, and developing positive attitudes. The following statements of teachers support this situation.

P11: We can attract students' attention to lessons with AI tools. In this way, we serve more effective learning.

P15: The current generation was born in the age of technology. It is almost impossible to keep them away from technology. Instead of keeping them away from children, we can use technology-supported AI tools to increase their motivation for lessons. These tools can motivate children to use creative solutions and make the lesson effective.

3.2.3. Other contributions

Teachers stated that AI tools can help students while doing homework and that students can use these tools in the communication process. In addition, according to teachers, advantages such as providing equal opportunities in crowded classes, reducing the education gap between villages and cities, and saving time and space are some of the positive contributions of AI in education. According to participant statements, it is also emphasized that AI provides rapid access to information, supports the education process with visual and audio materials, offers alternative learning methods, and increases the memorability of the acquired knowledge. These statements reflect a generally positive perspective that AI can offer many positive educational outcomes. Some of the teacher opinions supporting these statements are as follows;

P14: AI tools can serve individual learning in crowded classes."

P13: Using AI tools in education has advantages such as reducing the education gap between villages and cities and saving time and space. We will witness educational environments changing shortly.

P9: AI tools can save time and space by bringing learning environments into the classroom around the world.

P8: AI tools will serve to accelerate access to information. In particular, supporting the education process with AI tools and visual and audio materials will offer alternative learning methods and increase the memorability of the knowledge students acquire.

3.3. Theme 3: The role of AI in the educational environment

This theme focuses on the role of AI in educational environments. Teachers think that AI tools will make their jobs easier. Additionally, teachers believe that it is not possible for AI tools to replace teachers. According to teachers, the role of AI tools in the educational environment is summarized in Table 4 through themes, subthemes, and codes.

Table 4

The role of AI in the educational environment

Themes	Sub-themes	Codes	f
The Role of AI in the Educational Environment	AI makes the teacher's job easier	Makes teaching easier	12
		Teachers can monitor student's progress	10
		Reduces teacher workload	10
		Adds speed and practicality to the teacher	9
		It enriches learning	9
	AI cannot replace the teacher	Education is a human business	9
		Machines cannot touch people	7
		Education cannot be achieved with only cognitive knowledge.	7
		It mechanizes students	6
		AI cannot give emotional education	5
		Teaching is a matter of conscience	3

When Table 4 is examined, some teachers stated that they could replace teacher Al, while others said that they would not replace the teacher at all. According to teachers, the role of AI in educational environments occurs in two contexts. 1) AI makes the teacher's job more accessible, and 2) AI cannot replace the teacher.

3.3.1. AI makes the teacher's job easier

In the context of this sub-theme, teachers emphasized that AI tools facilitate teachers' teaching and learning tasks. According to teachers, AI can make it easier for the teacher to teach the lesson, allow the teacher to monitor the student's progress, reduce the teacher's workload, add speed and practicality to the teacher, and enrich children's learning. For example;

> *P1: Although Al cannot replace the teacher, it will make his job easier and provide effective teaching. Teachers can monitor students with AI. It will be easier to evaluate and follow up with children. It will also deepen children's learning.*

> *P8: "The teaching profession has undergone many transformations over the years. Technology-supported training is inevitable in our lives. "AI will be inevitable, and we must look for ways to use it more effectively.*

3.3.2. AI cannot replace teachers

According to teachers, digital technologies are becoming widespread today, and the education sector is also changing. Although some teachers stated that AI could replace the teacher with this change, others noted that the human factor is still needed in basic education (preschool and primary school education) and that it will be difficult for AI to replace the teacher. Teachers stated that cultural transfer and supporting students' socialization skills would be possible with teachers. For example;

> *P6: Although Al may replace the teacher in the future, I think that the teacher cannot be given up in order to transfer the culture and socialize the students in basic education.*

> *P13: The pandemic showed us that the lack of teachers in primary education prevented children's development of many skills. This will be the case in AI. Although it is a problematic possibility in primary education, it will replace the teacher in the upper grades.*

Teachers who stated that it could not replace the teacher stated that education is a human job, that students should be provided with cognitive knowledge and affective skills, and that education is a matter of conscience. Therefore, it is difficult for AI to replace the teacher. P6 emphasized his views on this issue: "The teacher is a biopsychosocial being, I do not think he can replace the teacher." In addition, the teachers' opinions stated below also support this situation.

> *P2: I do not think AI can replace the teacher. The teaching profession is more than just a job of transferring information. AI cannot undertake a profession that requires one-to-one communication and where emotions and behaviors are at the forefront. "Al will always ignore the conscientious side of a person.*

> *P4: It can be integrated into educational environments. "It may save time for teachers and students, but I do not think it can replace the teacher since learning environments are not based solely on information exchange.*

> *P5: Even if Al can make emotional analyses and understand the characteristics of the students, he cannot empathize and remains artificial. This does not work in young children." Therefore, AI cannot replace the teacher based on young children.*

> *P13: Just as all machines and robots reduce human power, it is also possible for this to happen in education. However, since the teacher is also someone who has an emotional side, I do not think that he can completely replace him in this aspect.*

4. Discussion

In this study, which aims to determine teachers' views on using AI tools in learning environments, teachers primarily focused on the adverse effects of AI tools. Most PST emphasize that AI technologies may hurt students' health and learning processes. These adverse effects are generally classified as thinking skills, socialization problems, and worthlessness. There are studies in the literature that support our findings. For example, Schiff (2020) noted concerns about using AI systems in education. One of these concerns is that AI systems could manipulate students by deeply analyzing them. Additionally, AI will likely standardize teachers, alienating them from their creativity and appreciation of students' diversity. A system that can understand all the characteristics of students has the potential to guide the students to make decisions on various issues. These concerns are essential factors preventing AI from fully integrating into teaching. Therefore, AI's role in education must be carefully considered, and appropriate measures must be taken (Akyel & Tur, 2024).

Since AI systems are based on extensive data analysis, they can sometimes produce incorrect results for specific cases (Chassignol et al., 2018). As Stephen Hawking and his colleagues noted in research published in The Independent in 2014, the short-term impact of AI may depend on who will use it and how it will benefit. However, in the long term, it is essential to determine whether the impact of AI can be controlled and whether individual and collective measures should be taken to prevent related risks (Hawking et al., 2014). Therefore, it is important to consider ethical concerns and ensure security during the development and use of AI (Yılmaz, 2023). Therefore, concerns about AI systems accelerate the taking of some precautions. Software developers who develop AI strive to reduce the margin of error and maintain ethical boundaries. For example, Microsoft introduced the "responsible AI" concept in 2017, identifying six ethical principles developers and institutions must meet: fairness, reliability and security, privacy, inclusivity, transparency, and accountability. Studies on ethical firsts can be shown as a promising effort in AI (Öztürk Dilek, 2019).

On the other hand, using artificial intelligence (AI) tools in education while offering innovative opportunities also brings various ethical concerns. The most important concerns are the privacy and security of students' data, AI systems' transparency, algorithms' unfairness, and the weakening of social interactions. In order to address privacy concerns, clear policies should be established regarding the data collection and use methods of AI tools, and these policies should be regularly audited. In order to ensure transparency, teachers, students, and parents should be informed about how AI systems work, and training should be organized. In addition, fair algorithms should be developed, data diversity should be ensured, and discrimination should be prevented. In order to prevent the weakening of social interactions, it is essential to balance the use of AI tools with activities that develop social skills. These solutions aim to ensure the ethical use of AI tools and to support students' academic and social development (Huang, 2023; Liang, 2023).

According to another finding, teachers also stated that there are advantages to using AI tools in education and training environments. These advantages are generally examined under three sub-themes: cognitive contribution, socio-emotional contribution, and others. Teachers think AI tools can facilitate access to information, support learning, and increase students' motivation. It has also been stated that artificial AI can assist students in the homework process and provide equal opportunities. AI is an information technology that can perceive human cognition, reason, comprehend, make sense of, generalize, infer, learn, and successfully perform multiple tasks simultaneously (Gondal, 2018), in a sense, imitating human Intelligence. Both participant opinions and related research indicate that using AI integrated with the teacher in all processes in the classroom and acting as an assistant will be more effective in classroom management (Bryant et al., 2020; Zhao & Lui, 2018). Studies conducted on AI tools in education Have shown that they can be used in grading and evaluation, student absenteeism and school dropout prediction, student performance prediction, personalized teaching, emotion analysis,

recommendation systems, innovative education systems, classroom monitoring, brilliant school, school evaluation and management, supervision, and analysis systems (Ahmad et al. al., 2020; Chen et al., 2020). AI technology can be used as a learning tool that provides practical learning experiences and supports learning by reducing the workload of both teachers and students in education. Studies are being carried out in this direction (Loeckx, 2016).

For example, AI contributes directly to school management and education, including course schedules, personnel schedules, exam management, cyber security, facility management, and security (Holmes et al., 2019). According to Edwards & Cheok (2018), if AI completes its development in educational presentation, personality, pedagogical approach, movement system, and emotional areas, another benefit will be that countries will eliminate the teacher shortage. AI is used in education as a support tool to automatically predict the development of students using their past performances and provide innovative teaching by improving education (Siegle, 2023). As a tool that contributes to creativity, AI provides advantages to teaching capabilities when used in art education (Caramiaux & Fidili Alaoui, 2022). While AI helps the student collaborate with technology, it places the learning process at the center of the student. It encourages him to take an active role by keeping him in a leadership position (Ezzaim et al., 2022). Teachers should determine a roadmap for students to use AI responsibly, safely, and effectively. This road map should include elements such as determining boundaries regarding AI, monitoring use, encouraging critical thinking, emphasizing responsible use, and providing guidance for the effective use of AI (Siegle, 2023).

Finally, in the evaluations made on the role of AI in educational environments, most teachers think that AI can make the teacher's job easier but cannot completely replace the teacher. In other words, teachers emphasized that education is human work, and the human factor is essential. According to teachers, although AI can partially replace the teacher in education, the teacher is the most critical factor in the education process. Therefore, AI will not be able to replace the teacher in primary education because basic education is a level where education is at the forefront rather than education. To motivate students to learn, the teacher is an integral part of the learning process and cannot be completely isolated from education. Some researchers support these findings in the literature. In a study, Kolchenko (2018) emphasizes that artificial intelligence will be insufficient in creating the templates formed in students due to student-teacher interaction. Therefore, the literature supports our results.

Similarly, Felix (2021) states that AI cannot teach cultural values, history, and social norms and that human teaching is necessary. He states that he is one step ahead with his posture, voice, gestures, and facial expressions and that AI cannot replace the teacher with its abilities, such as establishing eye contact, changes in tone of voice, and active use of the body. AI cannot imitate these features, so it is unlikely to displace the teacher. Bryant et al. (2020) state that having an assistant role would be more beneficial. The teacher will spend more time on education. Zhao & Liu (2018) state that AI will not replace the teacher in education but will remain an assistant to the teacher and state that education is a unique concept. It states that education is provided by transferring knowledge, developing character, discovering the unknown, and encouraging creativity. In this context, the literature supports our results.

5. Limitations and Recommendations

This study has some limitations. First, the number and profiles of teachers included in the study are limited. The experience and knowledge levels of the teachers participating in the study vary. This can lead to a need for more diversity and depth in the assessments and conclusions made. A study can use a larger sample, including different teacher profiles. Evaluating the opinions of students, parents, administrators, and teachers from different schools and regions can provide a more comprehensive perspective. A more detailed understanding of AI tools' potential effects and applications in education can be developed by examining teachers' concerns and suggestions.

Understanding the transformative potential of AI in education is a crucial starting point for this exploration. This understanding should delve into how AI-based tools can be seamlessly integrated into student learning processes, bolster teaching strategies, and foster personalized learning experiences. Equally important is the identification of the requisite support and training mechanisms for teachers to proficiently utilize these technologies. In this vein, regular professional development programs for teachers, along with guidelines on how to effectively align AI tools with pedagogical practices, and real-world examples of classroom applications, will all play a pivotal role in facilitating technology integration and enhancing the quality of education. A practical implementation strategy can then be devised to harness the full potential of these innovations and opportunities in education, ensuring that teachers perceive technology as a tool and an integral part of their teaching processes.

6. Conclusion

The results of this study broadly evaluate and discuss teachers' views on using AI tools in educational settings. Findings focus on the potential impacts of AI tools on students, their advantages, disadvantages, and their role in education. Teachers are concerned about the adverse effects of AI tools on students. These concerns are examined under themes such as the impact on thinking skills, socialization problems, and weakening of values. However, teachers also see the advantages of AI tools. These benefits include cognitive contribution, socio-emotional contribution, and other benefits. Finally, teachers also evaluated the role of AI in education. Most think AI can simplify the teacher's job but cannot completely replace it. In other words, AI can partially replace the teacher in the teaching process, but an education without a teacher is unthinkable in the education process, especially in primary education. These findings highlight the need to carefully consider the role of AI tools in education and the importance of creating appropriate guidance and policies for teachers to use AI effectively and responsibly. To maximize the potential of AI in education and minimize possible risks, ethical boundaries must be determined, and security measures must be taken. In this context, it is essential to conduct further research on AI tools' role in education and develop guidance materials that will enable teachers to use this technology effectively.

References

Ahmad, K., Iqbal, W., Hassan, A., Qadir, J., Benhaddou, D., Ayyash, M., & Fuqaha, A. (2020). *Data-Driven Artificial Intelligence in Education: A Comprehensive Review.* https://doi.org/10.35542/osf.io/zvu2n

Akgun, S., & Greenhow, C. (2022). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI and Ethics*, 2(3), 431-440. https://doi.org/10.1007/s43681-021-00096-7

Akyel, Y., & Tur, E. (2024). Yapay zekânın potansiyelinin ve eğitim bilimlerindeki uygulamalarının araştırılması ve araştırmalarda beklentiler, zorluklar ve gelecek yönelimleri [Exploring the Potential of Artificial Intelligence and Its Applications in Educational Sciences: Expectations, Challenges, and Future Directions in Research]. *Ahi Evran Üniversitesi Kırşehir Eğitim Fakültesi Dergisi,* 25(1), 645-712. https://doi.org/10.29299/kefad.1322341

Asimov, I. (2020). *I, Robot* (E. Odabaş, Çev.). İthaki Yayınları.

Ayala-Pazmiño, M. (2023). Inteligencia artificial en la educación: Explorando los beneficios y riesgos potenciales [Artificial Intelligence in Education: Exploring the Potential Benefits and Risks]. *593 Digital Publisher CEIT*, 8(3), 892-899. https://doi.org/10.33386/593dp.2023.3.1827

Bayraktar, B., Gülderen, S., Akça, S., & Serin, E. (2023). Yapay zekâ teknolojilerinin eğitimde kullanımına yönelik öğretmen görüşleri [Teachers' views on the use of artificial intelligence technologies in education]. *Ulusal Eğitim Dergisi*, 3(11), 2012–2030. Retrieved from https://uleder.com/index.php/uleder/article/view/380

Bowman, J. (2023). *How artificial intelligence is used in education, the motley fool.* (13 Nov 2023). Retrieved from https://www.fool.com/investing/stock-market/market-sectors/information-technology/ai-stocks/ai-in-education/

Bryant, J., Heitz, C., Sanghvi, S., & Wagle, D. (2020). *How artificial intelligence will impact K-12 teachers.* Retrieved from https://www.mckinsey.com/industries/education/our-insights/how-artificial-intelligence-will-impact-k-12-teachers

Bulut, M. A., Davarcı, M., Bozdoğan, N. K., & Sarpkaya, Y. (2024). Yapay zekânın eğitim üzerindeki etkileri [The effects of artificial intelligence on education]. *Ulusal Eğitim Dergisi*, 4(3), 976–986. Retrieved from https://uleder.com/index.php/uleder/article/view/498

Caramiaux B., & Fdili Alaoui S. (2022). "*Explorers of unknown planets": Practices and politics of artificial ıntelligence in visual arts*. Conference On Computer-Supported Cooperative Work and Social Computing, November 12–16. https://doi.org/10.1145/3555578

Celik, I., Dindar, M., Muukkonen, H., & Järvelä, S. (2022). The promises and challenges of artificial intelligence for teachers: A systematic review of research. *TechTrends,* 66(4), 616-630. https://doi.org/10.1007/s11528-022-00715-y

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, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access,* 8, 75264 – 75278. https://doi.org/10.1109/access.2020.2988510

Chou, J., Murillo, O., & Ibars, R. (26 Sep, 2017). How to Recognize Exclusion in AI. *Medium.* Retrieved from https://medium.com/microsoft-design/how-to-recognize-exclusion-in-ai-ec2d6d89f850

Druga, S., & Ko, A. J. (2021, June). How do children's perceptions of machine intelligence change when training and coding smart programs?. In *Proceedings of the 20th annual ACM interaction design and children conference* (pp. 49-61).

Edwards, B. I., & Cheok, A. D. (2018). Why not robot teachers: Artificial intelligence for addressing teacher shortage. *Applied Artificial Intelligence,* 32(4), 345-360. https://doi.org/10.1080/08839514.2018.1464286

Edwards, S. (2023). Early childhood education and care, artificial intelligence and the post digital. *Australasian Journal of Early Childhood*, 48(3), 179-181. https://doi.org/10.1177/18369391231198910

Eren, Z. (2021). Eğitimde yapay zekâ uygulamaları ve geleceğe ilişkin yönelimler [Artificial Intelligence Applications in Education and Future Trends]. (Eds. N. Öykü İyigün & Mustafa K. Yılmaz). *Yapay zekâ: güncel yaklaşımlar ve uygulamalar [Artificial Intelligence: Current Approaches and Applications].* İstanbul: Beta Kitap Yayıncılık.

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

Felix, C. V. (2021). The role of the teacher and AI in education. Sengupta, E., Blessinger, P. and Makhanya, M.S. (Der.). *International perspectives on the role of technology in humanizing higher education* (Innovations in higher education teaching and learning, 33), In (pp.33-48). Emerald Publishing Limited. https://doi.org/10.1108/S2055-364120200000033003

Gondal, K. M. (2018). Artificial intelligence and educational leadership. *Annals of King Edward Medical University,* 24(4), 1-2. https://doi.org/10.21649/akemu.v24i4.2591

Gordon, B. M. (2011). *Artificial intelligence: Approaches, tools and applications.* New York: Nova Science Publishers, Inc.

Hawking, S., Russel, S., Tegmark, M., & Wilczek, F., (2014). *Transcendence looks at the implications of artificial intelligence- but are we taking AI seriously enough*? The Independent. Retrieved from https://www.independent.co.uk/news/science/stephen-hawking-transcendence-looks-at-the-implications-of-artificial-intelligence-but-are-we-taking-ai-seriously-enough-9313474.html

Holmes W., Persson J., Chounta I. A., Wasson B., & Dimitrova V. (2022). *Artificial intelligence and education: A critical view through the lens of human rights, democracy, and the rule of law*. Council of Europe.

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning.* Boston, MA: Center for Curriculum Redesign.

Kavale, K. A., & Forness, S. R. (2019). *Cheating and plagiarism in schools and colleges*. Psychology Press

Kim, S., Jang, Y., Kim, W., Choi, S., Jung, H., Kim, S., & Kim, H. (2021, May). Why and what to teach: AI curriculum for elementary school. In *proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 35, No. 17, pp. 15569-15576).

Knox J. (2019). What does the 'postdigital' mean for education? Three critical perspectives on the digital, with implications for educational research and practice. *Postdigital Science and Education*, 1(1), 357–370. https://doi.org/10.1007/s42438-019-00045-y

Kolchenko, V. (2018). Can modern AI replace teachers? Not so fast! Artificial intelligence and adaptive learning: Personalized education in the AI age. *HAPS Educator,* 22(3), 249-252. https://doi.org/10.21692/haps.2018.032

Kurian, N. (2023). AI's empathy gap: The risks of conversational Artificial Intelligence for young children's well-being and key ethical considerations for early childhood education and care. *Contemporary Issues in Early Childhood*, *(early view),* 14639491231206004. https://doi.org/10.1177/14639491231206004

Liang, J. C., Hwang, G. J., Chen, M. R. A., & Darmawansah, D. (2023). Roles and research foci of artificial intelligence in language education: An integrated bibliographic analysis and sysThemestic review approach. *Interactive Learning Environments*, 31(7), 4270-4296. https://doi.org/10.1080/10494820.2021.1958348

Lin, P., Van Brummelen, J., Lukin, G., Williams, R., & Breazeal, C. (2020). Zhorai: Designing a Conversational Agent for Children to Explore Machine Learning Concepts. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(09), 13381-13388. https://doi.org/10.1609/aaai.v34i09.7061

Loeckx, J. (2016), Blurring boundaries in education: Conext and impact of MOOCs. *International Review of Research in Open and Distributed Learning,* 17(3), 93-121. https://doi.org/10.19173/irrodl.v17i3.2395

Luckin, R., & Cukurova, M. (2019). Designing educational technologies in the age of AI: A learning sciences-driven approach. *British Journal of Educational Technology*, 50(6), 2824–2838. https://doi.org/10.1111/bjet.12861

McCarthy, J. (2007). From here to human-level AI. *Artificial Intelligence*, 171(18), 1174-1182. https://doi.org/10.1016/j.artint.2007.10.009

Mengesha Z, Heldreth C, Lahav M, et al. (2021). "I don't think these devices are very culturally sensitive." Impact of automated speech recognition errors on African Americans. *Frontiers in Artificial Intelligence*, 26(4), 1-12. https://doi.org/10.3389/frai.2021.725911

Merriam, S. B. (2015). *Nitel araştırma desen ve uygulama için bir rehber [Qualitative Research: A Guide for Design and Implementation]*. (Trans. Eds. Turan, S.). Nobel Akademi Yayıncılık.

Minkkinen, M., & Mäntymäki, M. (2023). Discerning between the "easy" and "hard" problems of AI governance. *IEEE Transactions on Technology and Society*, 4(2), 188–194. https://doi.org/10.1109/TTS.2023.3267382

Mondal, B. (2020). Artificial intelligence: state of the art. In: Balas, V., Kumar, R., Srivastava, R. (eds) *Recent trends and advances in artificial intelligence and internet of things*. Intelligent Systems Reference Library, 389-425. Springer, Cham. https://doi.org/10.1007/978-3-030-32644-9_32

Öztürk Dilek, G. (2019). Yapay zekanin etik gerçekliği [The Ethical Reality of Artificial Intelligence]. *Ankara Uluslararası Sosyal Bilimler Dergisi,* 2(4), 47-59. Retrieved from https://dergipark.org.tr/en/pub/usdad/issue/51335/642184

Patton, M.Q. (2015). *Nitel araştırma ve değerlendirme yöntemleri [Qualitative Research and Evaluation Methods].* (Trans. Eds. Bütün, M., & Demir, S. B.). Pegem akademi yayıncılık: Ankara.

Qin, F., Li, K., & Yan, J. (2020). Understanding user trust in artificial intelligence based educational systems: Evidence from China. *British Journal of Educational Technology,* 51(5), 1693-1710. https://doi.org/10.1111/bjet.12994

Schiff, D. (2020). *Out of the laboratory and into the classroom: The future of artificial intelligence in education. AI and Society*. Springer. https://doi.org/10.1007/s00146-020-01033-8

Sevil, Ş., & Saralar-Aras, İ. (2024). *Eğitimde kullanılan yapay zeka araçları öğretmen el kitabı [Teacher's Handbook on Artificial Intelligence Tools in Education]*. Ankara. Retrieved from https://yegitek.meb.gov.tr/www/egitimde-kullanilan-yapay-zek-araclari-ogretmen-el-kitabi-yayimlandi/icerik/3631

Seyrek, M., Yıldız, S., Emeksiz, H., Şahin, A., & Türkmen, M. T. (2024). Öğretmenlerin eğitimde yapay zekâ kullanımına yönelik algıları [Teachers' Perceptions of Artificial Intelligence Use in Education]. *International Journal of Social and Humanities Sciences Research (JSHSR)*, 11(106), 845–856. https://doi.org/10.5281/zenodo.11113077

Siegle, D. (2023). A role for chatgpt and AI in gifted education. *Gifted Child Today,* 46(3), 211-219. https://doi.org/10.1177/10762175231168443

Su, J., & Yang, W. (2022). Artificial intelligence in early childhood education: A scoping review. *Computers and Education: Artificial Intelligence*, 3, 100049. https://doi.org/10.1016/j.caeai.2022.100049

Su, J., & Zhong, Y. (2022). Artificial Intelligence (AI) in early childhood education: Curriculum design and future directions. *Computers and Education: Artificial Intelligence*, 3, 100072. https://doi.org/10.1016/j.caeai.2022.100072

Su, J., Ng, D. T. K., & Chu, S. K. W. (2023). Artificial intelligence (AI) literacy in early childhood education: The challenges and opportunities. *Computers and Education: Artificial Intelligence*, 4, 100124. https://doi.org/10.1016/j.caeai.2023.100124

Tseng, T., Murai, Y., Freed, N., Gelosi, D., Ta, T. D., & Kawahara, Y. (2021). PlushPal: Storytelling with interactive plush toys and machine learning. In *Interaction design and children* (pp. 236-245).

UNİCEF, (2021). Children and AI Where are the opportunities and risks. Retrieved from *Unicef. https://www.unicef.org/innovation/sites/unicef.org.innovation/files/2018-11/Children%20and%20AI_Short%20Verson*, 20(283), 29.

Williams, R., Park, H. W., Oh, L., & Breazeal, C. (2019). Popbots: Designing an artificial intelligence curriculum for early childhood education. *Proceedings of the AAAI Conference on Artificial Intelligence,* 33(1), 9729–9736.

Yang, W. (2022). Artificial Intelligence education for young children: Why, what, and how in curriculum design and implementation. *Computers & Education: Artificial Intelligence*, 100061. https://doi.org/10.1016/j.caeai.2022.100061

Yi, H., Liu, T., & Lan, G. (2024). The key artificial intelligence technologies in early childhood education: a review. *Artificial Intelligence Review*, 57(1), 12. https://doi.org/10.1007/s10462-023-10637-7

Yıldırım, H., & Şimşek, A. (2013). *Sosyal bilimlerde nitel araştırma yöntemleri [Qualitative Research Methods in Social Sciences]*. Seçkin Yayıncılık.

Yılmaz, Z. (2023). "Yapay zeka ve felsefe" soruşturması [Inquiry into "Artificial Intelligence and Philosophy"]. *Düşünce ve Toplum Sosyal Bilimler Dergisi,* 5(2), 159-201. https://doi.org/10.55796/dusuncevetoplum.1412018

Zanetti, M., Rendina, S., Piceci, L., & Cassese, F. P. (2020). Potential risks of artificial intelligence in education. *Form@ re-Open Journal per la Formazione in Rete*, 20(1), 368-378. https://doi.org/10.13128/form-8113

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators?. *International Journal of Educational Technology in Higher Education*, 16(1), 1-27. https://doi.org/10.1186/s41239-019-0171-0

Zhang, K., & Aslan, A. B. (2021). AI technologies for education: Recent research & future directions. *Computers and Education: Artificial Intelligence*, 2, 100025. https://doi.org/10.1016/j.caeai.2021.100025

Zhang, P., & Tur, G. (2024). A systematic review of ChatGPT use in K-12 education. *European Journal of Education*, 59(2), e12599. https://doi.org/10.1111/ejed.12599

Zhao, Y., & Liu, G. (2018). How do teachers face educational changes in artificial intelligence era. *Advances in Social Science, Education and Humanities Research (ASSEHR),* 300, 47-50. Retrieved from https://www.atlantis-press.com/proceedings/erss-18/55912681

Zheng, L., Niu, J., Zhong, L., & Gyasi, J. F. (2023). The effectiveness of artificial intelligence on learning achievement and learning perception: A meta-analysis. *Interactive Learning Environments*, 31(9), 5650-5664. https://doi.org/10.1080/10494820.2021.2015693

Liang, Y. (2023). Balancing: The Effects of AI Tools in Educational Context. *Frontiers in Humanities and Social Sciences*, 3(8), 7-10. https://doi.org/10.54691/fhss.v3i8.5531

Huang, L. (2023). Ethics of artificial intelligence in education: Student privacy and data protection. *Science Insights Education Frontiers,* 16(2), 2577-2587. https://doi.org/10.15354/sief.23.re202

Article Information Form

The Adventure of Artificial Intelligence in Educational Research from the Past to the Present

Deniz Kaya ⓘD

Nevşehir Hacı Bektaş Veli University,
Faculty of Education, Nevşehir, Türkiye,
denizkaya@nevsehir.edu.tr

Abstract: This study aims to examine scientific studies on artificial intelligence (AI) in educational research from the past to the present, based on the Web of Science database. In this context, 1465 scientific articles containing AI in education from the past to the present were evaluated. Articles accessed from the WoS database were examined using a bibliometric analysis method according to productivity, network analyses, conceptual structure, and thematic map titles. Within the scope of productivity, authors, institutions, countries, citations within the scope of network analysis, authors, institutions, sources, and countries were included in the analysis. In addition, thematic changes over the years, word cloud, collaborations, conceptual formations, and thematic mapping were carried out based on keywords. In this context, 1465 scientific articles published by 3783 authors representing 86 countries were included in the research. According to the research findings, the number of studies and citations on AI in education has increased significantly, especially in the last five years. The Education University and The Chinese University of Hong Kong stand out as productive institutions. While China, England, and the USA stand out as the countries of responsible authors, Hwang, G. J., stands out as the author of network analysis, and the Computer Education journal stands out as the journal. As a thematic change in the studies, there has been an evolution towards new technological developments such as deep learning, machine learning, ChatGPT, chatbots, learning analytics, blockchain, and generative AI. According to the factor analysis conducted on the conceptual structure of AI-related studies in education, it was determined that it explained 48% of the total variability. According to the study findings, studies on AI applications in education should be enriched from a disciplinary perspective, and efficiency should be increased regarding their reflections on teaching.

Keywords: Artificial Intelligence (AI), Bibliometric Analysis, Educational Research, Web of Science (WoS)

1. Introduction

Innovative technological research is undoubtedly one of the main components that make countries stand out in global competition. In this regard, strengthening scientific and technological capacity and sustaining the technology ecosystem has become inevitable for many nations. With the technological developments in recent years, concepts such as artificial intelligence (AI), internet of things (IoT), robotic coding, blockchain, metaverse, big data, nanotechnology, digital change, and virtual reality have become increasingly prevalent in our lives, and there is an increasing interest in these areas. Changes that require AI technologies, especially voice recognition, facial recognition, and autonomous vehicles, are significantly affecting societies' lives (Huang & Qiao, 2024). The rapid advancement of technology causes changes in habits on a global scale and the differentiations of ways of connecting, interacting, reading, writing, and being informed (Hinojo-Lucena et al., 2019). So much so that, thanks to the successful applications of AI, there is a sharp evolution towards adaptive intelligence software, and the application areas of innovative products are diversifying at a rapid pace, demonstrating the breadth of its impact. As in many disciplines, the effects of innovative technological concepts are strongly felt in education. Especially under AI, virtual digitalization, big data, and IoT, encouraging the speed of modernization of the ecological structure of education systems comes to the fore (Huang et al., 2021a). With a better understanding of the potential capabilities of productive AI, individuals' learning knowledge is expanded by producing quality content. For this reason, there is a sharp evolution towards an AI society that individuals can easily experience anytime and anywhere. We are witnessing the inevitable rise of AI applications in many areas, such as personalized online education systems, medical

services, agriculture, manufacturing, communication, media, transportation, defense, communications, logistics, weather (Organisation for Economic Co-operation and Development [OECD], 2019). In education, as in many fields, the effects of AI-based applications are felt better and included in learning environments with innovative applications.

Today, the necessity of nations redesigning their education policies is increasing due to the values that AI-based technological transformations add to learning. It has been observed that AI-based new-generation technologies in education provide successful pedagogical results with the help of various applications such as content presentation, feedback, and progress control with an intelligent teacher (Bayne, 2015; Chen et al., 2022). For this reason, AI, a machine-based technique with algorithmic power, has been frequently included in education in recent years to support learning in various contexts (Hwang et al., 2020). There is increasing focus on applying AI technology to create intelligent campuses, assist education, and efficiently carry out learning by producing intelligent learning-teach algorithms (Huang et al., 2021b). Therefore, it has been essential to understand AI-related issues in education over the years, determine general trends, and provide suggestions for future researchers.

1.1. Conceptual framework

1.1.1. Artificial intelligence

AI is a new-generation technology product that refers to imitating basic abilities such as thinking, learning, and decision-making that distinguish humans from other living things through electronic devices, especially computers. AI, also called machine intelligence, is a sub-branch of computer science that focuses on producing a new type of intelligent machine that simulates human intelligence (Huang & Qiao, 2024). The concept of AI is computer systems that can perform human-specific processes such as learning, adaptation, synthesis, self-correction, and use of data for complex processing tasks. (Popenici & Kerr, 2017). In other words, a processor with tremendous capabilities includes adaptive behaviors and human-like cognitive and functional abilities (Chen et al., 2020). AI, a strategic technology, is pioneering a new era in technological, industrial, or social fields, creating significant and far-reaching effects on education, economic, and social situations (United Nations Educational, Scientific and Cultural Organization [UNESCO], 2021). The proposed concept of artificial neural networks fueled the origin of these effects. Artificial neural networks were designed by McCulloch and Pitts in 1943 as an algorithmic model that took the functionality of the human brain into account and was organized as a system structure consisting of a large number of neurons connected in parallel and interacting with real-world examples (Kohonen, 2001). In this regard, the development of AI is not problem-free, and the lack of methodology is always felt due to widespread expectations.

AI permeates every layer of our lives and is reflected in many disciplines. Considering that AI and human intelligence coexist, there will not be a field of work in the future where the impact of AI is not felt and its impact is not mentioned. AI, which has a versatile impact on all areas, including purpose, content, method, and evaluation system, also has multidimensional effects on education (Paek & Kim, 2021). Today, AI applications solve real-world problems in six areas: computer vision, machine learning, natural language processing, cognition and reasoning, robotics, games, and ethics (Huang & Qiao, 2024). In this regard, AI is built on three growth factors: algorithms, big data, and computing power. Especially in the early 21st century, the increase in data volume with the widespread use of smartphone applications and the diversification of the number of tools with algorithm extensions has significantly contributed to the development of AI (Russell & Norvig, 2021). Thanks to machine learning and deep learning breakthroughs in AI-based technologies, as structures containing big data, cloud computing, and related computing and storage performance become available, AI's performance and impact areas increase daily (Chatti et al., 2012; Garcia et al., 2007; OECD, 2019).

1.1.2. Artificial intelligence in education

In today's information age, the first quarter of which we are only experiencing, radical reforms are being experienced in education with the rapid development of AI technologies (Huang et al., 2021a, 2021b). In particular, the fact that AI applications require a multifaceted disciplinary perspective, such as science, mathematics, sociology, psychology, engineering, philosophy, and geography, accelerates the transformation of these reforms. AI applications that impact many areas of our lives are frequently encountered in education (Bozkurt et al., 2021). Among the primary reasons for using AI in educational environments is to produce effective, qualified, high-quality, and fast solutions to problems encountered in daily life. For this reason, personalized systems, software, ontologies, and semantic web techniques stand out as areas of use for AI in education (Lemaignan et al., 2017). According to Yang and Zhang (2019), AI in intelligent teaching systems can determine learning performance, knowledge level, intelligence level, preferences, learning style, learning behavior, and cognitive, affective, and cultural factors. Therefore, the reflections of AI applications on education can be evaluated multifacetedly. Drawing attention to this situation, Chen et al. (2020) state that AI significantly impacts the education sector, especially in management, teaching, and learning. This is because AI can be organized as a learning material according to student needs, enables personal learning, and has the potential for a better learning experience. In addition, AI-supported educational environments allow for analyzing student participation, identifying at-risk students full-time, and shortening the intervention time in the learning environment (Chen et al., 2022; Hwang et al., 2020; Tsai et al., 2020).

Today, AI applications in education are accepted in many areas. AI applications are increasing in popularity today in order to promote fair and qualified progress in education and to determine their benefits for teachers and learning (Wang et al., 2023). Creating applications that increase individual learning capacity, developing personalized learning resources, and enriching the perception of learning with simulated scenarios can be cited as the contributions of AI to education (Shi et al., 2024; Xie et al., 2019). Many researchers; believe that the use of AI in education will improve their students' digital literacy, knowledge, collaboration, learning abilities, and academic perceptions (Huang, 2021; Lee & Lee, 2021; Wang et al., 2023). It includes improving the preferred AI learning effects and teaching mode in education and basic AI knowledge and how it will interact (Huang & Qiao, 2024). In this context, numerous research outputs focusing on the application of AI in education suggest that AI will become a technical tool with much stronger potential for future learners and educators (Chatti et al., 2012; Garcia et al., 2007; Paek & Kim, 2021; Wang et al., 2023). Image and face recognition, adaptive learning, and other AI-based technological applications provide a learning experience, increase efficiency, and offer a different educational perspective (Cui et al., 2018; Hwang & Tu, 2021). For this reason, the role of AI applications in education is increasing daily.

1.1.3. Literature review on artificial intelligence in educational research

Although AI first emerged in the mid-20th century, studies/research in this field have recently gained significant momentum. Although there is increasing interest in AI and its applications, more studies need to be reviewed to investigate the use of AI in education. When research on AI in education is generally evaluated, it can be seen that issues such as the development of computational thinking skills through AI (Huang & Qiao, 2024), curriculum design based on AI (Chiu & Chai, 2020), the potential of sports applications in physical education (Lee & Lee, 2021), nursing pain education (Harmon et al., 2021), the role of AI in mathematics education (Hwang & Tu, 2021), the association of AI and virtual reality (Lin et al., 2021), AI and flipped learning-based mental health education (Shan & Liu, 2021), students' flow experience and learning effectiveness (Shi et al., 2024), intelligent homework grading and brilliant question answering systems (Tobler, 2024), offline course applications (Li & Wang, 2021), and ethics in the use of AI (UNESCO, 2021) are discussed. Common indicators of these studies include the fact that productive AI-based applications positively affect students' learning intentions (Lin et al., 2021;

Shan & Liu, 2021), promote the modernization of education by providing enriched learning resources (Zhao et al., 2023), improve students' performance by attracting their attention (Li & Wang, 2021; Shan & Liu, 2021; Wang et al., 2023), enhance creativity, offer a live learning experience, increase self-efficacy and improve digital literacy (Garcia et al., 2007; Wang et al., 2023).

We see evaluations from different perspectives when we look at similar studies in the literature. For example, Bozkurt et al. (2021) systematically examined AI studies in education in half a century (1970-2020). According to the study's findings, there has been a significant increase in the number of studies involving AI in recent years, and this trend will probably increase in the coming years. Data obtained from WoS and Scopus databases were used in the bibliometric analysis study conducted by Hinojo-Lucena et al. (2019). According to the findings, AI studies need to be at a sufficient level to allow the production of scientific content. It has been stated that the USA stands out in productivity with the University of Alicante, Polytechnic University of Valencia, Autonomous University of Madrid, and University of Alcala institutions. In the bibliometric and content analysis study conducted by Bahroun et al. (2023), 207 research articles were examined. At the end of the study, it was reported that ChatGPT has emerged as a dominant generative AI tool and that there is an exponential increase in generative AI. The systematic review conducted by Forero-Corba and Negre-Bennasar (2024) examined studies on machine learning and AI. Fifty-five articles obtained from WoS and Scopus databases were examined. According to the findings, using machine learning and AI has strong effects. As a result of the study conducted by Paek and Kim (2021) based on the WoS database, it was determined that AI studies have increased dramatically in the last 20 years. At the same time, it has been reported that the issues related to AI education technology and measurement and evaluation are up-to-date. Chen et al. (2022) tried to determine the trends and topics related to AI by examining 4519 publications between 2000 and 2019. According to the research findings, it has been stated that the interest in using AI for educational purposes has increased, and the subjects of intelligent lesson systems, language education systems, educational robots, educational data mining for performance prediction, discourse analysis in collaborative learning, and teaching evaluation have come to the fore.

1.2. Purpose of the study

Technological developments deeply affect nations' perspectives on education and training. In this regard, having the competencies required by the information age and keeping up with innovative changes has become inevitable. Considering the increasing interest in AI-based applications, especially in the last decade, the combination of education and technology is felt more intensely. Therefore, it is essential to know the reflections of the studies on AI applications, which have influenced education worldwide. Future research is essential to holistically reveal AI-related studies' trends, impact, and potential, especially in the relevant literature. In this regard, the bibliometric analysis approach, which allows us to examine all aspects of the studies from past to present, helps us. Bibliometric-based research includes much quantitative information such as subject headings, contents, keywords, publication language, author, co-authors, authors' institution, authors' countries, reference information, reference impact levels, citations, co-citations, and year of publication. This type of research contributes to obtaining scientific findings with the help of quantitative analysis on the determined subject and allows the application of many different quantitative analysis methods. Bibliometrics supports statistically presenting information about the subject of study. It helps to present the importance of the subject in the literature to the readers quantitatively (Zhao & Strotmann, 2015). Such research provides valuable data to field experts, readers, and program makers by evaluating scientific content from many aspects. Given all the explanations, the study's starting point is to determine the general change, development, direction, and impact of AI-based studies in educational research. It is essential to know the general structure of the research topic in the relevant field and to reveal the needs by following it periodically. Therefore, it is hoped that this comprehensive research will constitute an essential

resource in the field, guide future research, and give ideas to experts in different disciplines. In addition, it is expected to accelerate research in similar directions and allow us to see the changes in the process better. This study evaluated articles containing AI in educational research indexed in the WoS database according to performance-based descriptive findings, network analyses, conceptual structure, and thematic mapping, and answers to the following research questions (RQs) were sought:

RQ 1. What is the change in AI-related research in education according to years and citation numbers?
RQ 2. What are the authors, resources, institutions, and countries that contribute to AI-related research in education?
RQ 3. Which authors and sources interact in AI-related research in education?
RQ 4. How does the collaborative profile of authors, institutions, and countries regarding AI-related research in education change?
RQ 5. How do common keyword and co-occurrence profiles change in AI-related research in education?
RQ 6. What are the trending topics and thematic changes in AI-related research in education?
RQ 7. How do the general conceptual structure and thematic mapping change in AI-related research in education?

2. Methodology

2.1. Research design

In this research, scientific articles published on AI from the past to the present were examined in a descriptive and cross-sectional retrospective manner with the help of bibliometric analysis. Bibliometric analysis was used in the research, making it possible to examine the scientific literature and researchers who contribute to it according to statistical procedures. This analysis provides a holistic perspective by creating a road map for readers and researchers regarding the determined research topic (Chen et al., 2019). In this analysis, applications are made according to the criteria determined by the data set created using quantitative techniques (Pritchard, 1969). One of the purposes of bibliometric analysis is to summarize an extensive data set in the context of specific criteria (Donthu et al., 2021). These analyses involve collecting, processing, and evaluating publications with scientific content (Verbeek et al., 2002). Bibliometrics is a practical application, especially in better defining and distinguishing the field (Donthu et al., 2021). Bibliometric analysis involves dynamics and structure (Chaparro & Rojas-Galeano, 2021). In dynamic analysis, indicators such as publications, citations, authors, keywords, and terms are examined within the scope of the scientific production network, and in structure analysis, indicators such as conceptual, interaction network, thematic change, and network are examined (Jamali et al., 2022). This analysis technique allows the evaluation of research results and examines scientific outputs comprehensively (Grzybowska & Awasthi, 2020). This study conducted bibliometric analyses according to performance (scientific productivity), network analysis, conceptual structure, and thematic mapping. In this regard, according to the research framework proposed by Arksey and O'Malley (2005), first, the research question was defined, the database was decided to create the data set related to the research topic, the database was accessed, the criteria to be used in data selection were determined, and a data set was created from the database, the data set was analyzed, the analyzes were evaluated, the findings were reported, and the processes were reviewed by providing interpretation of the findings. Process information, including the basic framework of the research, is given below (Figure 1).

Figure 1

The Basic Framework of the Research

2.2. Data collection and procedure

Web of Science™ Core Collection database was used to create the data set of educational studies with AI content. The Web of Science (WoS) database contains research categories with many options for analyzing published documents. This database includes many disciplines and provides reliable data with advanced search options. It also allows statistical analysis and quantitative techniques suitable for bibliometric analysis. The strengths of the WoS™ database include being easy to access, containing prestigious journals, being suitable for downloading appropriate data from analysis programs, allowing detailed searches in categories, being compatible with the programming language, containing reference information, and having an open-source representation (Web of Science Group [WoSG], 2024). The WoS™ database has been designed using a structure within Clarivate Analytics. Its many working areas have gained a respected place worldwide, allowing researchers to examine and record in-depth (Fang et al., 2017). To create the data set of the research, the Topic (search title, abstract, and author keywords) module was considered. Accordingly, a search was made in the WoS™ database for scientific records related to research on AI with educational content. The process followed to obtain scientific records is as follows: WoS™ Database: *[TITLE-ABS-KEY ("artificial intelligence") AND (educat* OR learn* OR teach* OR class* OR innovat* OR student*) Refined by: Document Type: (Article), Language: (English) and Years of Publication: All years, and Science Categories: (Education Educational Research) and Web of Science Index: (Social Science Citation Index (SSCI), Science Citation Index-Expanded (SCI-Expanded), Emerging Sources Citations Index (ESCI), Arts & Humanities Citation Index (A&HCI)].* The figure reflecting the screening criteria performed to create the data set is shown below (Figure 2).

Figure 2

Information Including the Criteria of the Screening
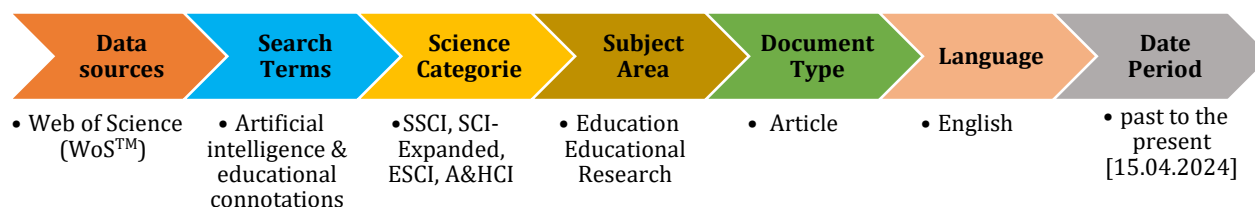
Documents downloaded from the WoS™ database are saved in the specified folder in "plain text" format. This file format works and is compatible with VOSviewer and RStudio applications. The WoS™ database allows downloading up to 1000 documents. These floating two separate files were downloaded and merged into one. This final data set was used for bibliometric analysis.

2.3. Data analysis process

Data obtained from the WoS™ Core Collection database was analyzed bibliometrically. The preferred bibliometric analysis application for the research topic consists of four stages. In the first stage, the data

set intended for research was accessed from the database. When the selection was made according to the criteria [*WoS™ Index: SSCI, SCI-Expanded, ESCI, A&HCI; Subject Area: Education & Educational Research; Document Type: Articles, Language: English; Publication Period: Past to the Present (15.04.2024)],* it was determined that there were 1465 scientific documents. At this stage, the data was checked and examined to determine whether there was any duplicate data. In the second stage, performance-based analyses were carried out. In this context, the annual number of publications and citations, the most published authors, the most cited studies, the distribution of responsible authors by country, the distribution of dominant authors by year, and the institutions with the most publications were examined. In the third stage, network analyses (analyses were used to include cited common citation, common author, geographical atlas, common word, trend topic, and thematic changes) were used. In the last stage, analyses were made according to conceptual formations and thematic mapping. The general structure, including the analysis, is presented below (Figure 3).

Figure 3

General Structure Including Analysis
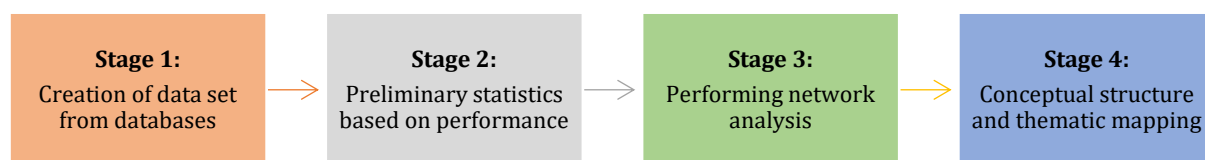
VOSviewer 1.6.18 software was used to analyze research data, which allows dynamic and structured analyses of large volumes of data (Van Eck & Waltman, 2010). The VOSviewer program uses labels and circular structures when visualizing network analyses. The weighting of the element, depending on the volume of the circle size, is calculated with the help of different colors, and similar elements are divided into clusters (Yuan et al., 2021). The amount of interaction depends on the intensity of the relationship between each cluster. The R programming language, which allows the data set to be examined from many aspects, was used to investigate conceptual and thematic changes. Thanks to this programming language, many contents can be analyzed, such as dominant authors by year, conceptual formations, thematic changes, geographical view word formations, transformational differences, and formations between countries related to the research subject (Aria & Cuccurullo, 2017). VOSviewer [https://www.vosviewer.com] and R-tool [www.rstudio.com], which are open-access and free applications that are frequently preferred in bibliometrics-related research, allow both in-depth and visualization of the given data (Van Eck & Waltman, 2010). Information containing a comprehensive view of the dataset using RStudio (biblioshiny) is presented in Figure 4.

Figure 4

A Comprehensive View of the Dataset
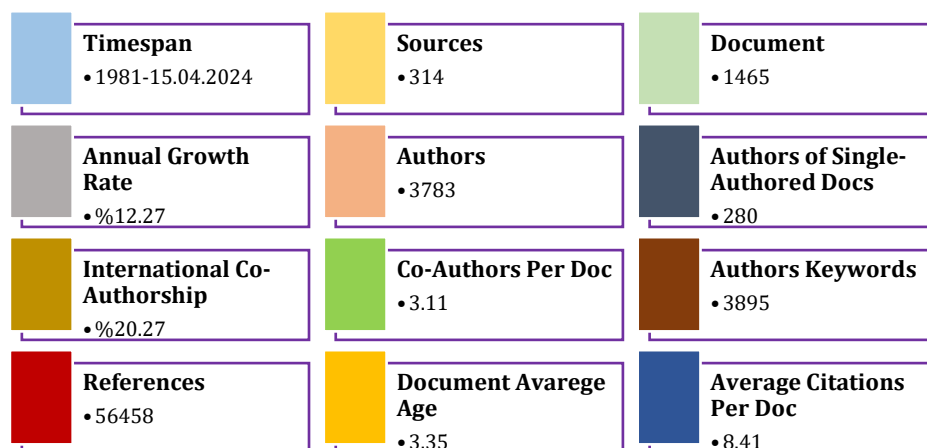
As shown in Figure 4, 1465 scientific articles containing AI in education were written by 3783 authors, according to the criteria determined from 1981 to the present [15.04.2024]. The number of scientific articles with a single author was 280, the collaboration index between authors was 3.11, the international collaboration index was 20.27%, the annual growth rate was 12.27%, the average number of citations was 8.41, the number of references was 56458 and document average age was 3.35. The thematic and strategic diagram technique in analyzing selected scientific studies creates dynamic clusters by analyzing key or co-axial words (Law et al., 1988). These clusters provide information about the general view of the research topic (Gonzales-Valiente, 2019). On the other hand, conceptual maps created regarding the determined research topic divide the content of the total data set into information sets and produce comprehensive results about the research content and trend (Wetzstein et al., 2019). This way, qualified findings are obtained to better organize future research areas and reveal different research patterns. In this context, the research analyzed data holistically by considering descriptive, social network, conceptual structure, and thematic changes.

2.4. Validity and reliability

Reliability and validity studies are among the basic requirements of research. First of all, to increase the research's validity, care was taken to express the actions taken regarding the research data process in detail. In this regard, it is clearly stated which database the data will be obtained from, the website of the database, the date on which the data was collected, what criteria were used when searching for the data in the database, the keywords used in the data search, what restrictions were made in the search module and in what format the data was recorded. For the reliability of external validity, the steps taken in analyzing the data were specified, and information was given about the analysis programs applied. In addition, the justification of the methods used in data analysis and the factors taken into account in forming the data set are explained together with their reasons. Further contributing to reliability, the findings were presented directly, and it was ensured that the findings were holistic. In addition, the results obtained were supported by the relevant literature to increase consistency. In addition, information about the analysis programs used in the research and the web addresses where this software can be accessed are stated.

3. Findings

In this part of the study, the findings of AI-related research in the field of education, depending on the determined sub-problems, are presented with their explanations. First, under scientific productivity, AI-related research in education is distributed according to years, the number of citations, and the results for the most contributing authors, studies, institutions, and countries. Immediately afterward, findings from co-citation and co-author analyses are presented. Another title presents collaboration networks between institutions and countries and word cloud analyses under keyword and co-occurrence analysis. In addition, trending keywords and article titles of AI-related research in the field of education were analyzed under the title of trend topic and thematic change. Finally, the conceptual formations of published articles containing artificial intelligence in education are reported.

3.1. Scientific productivity on ai in educational research

The study reports the adventure of AI-related studies in education, according to the WoS database, from 1981 to the present, according to one year. In addition to the annual number of articles and citations in different colors, the harmonic average of the citation numbers is also included. In this context, the findings obtained according to the number of articles and citations registered in the WoS database from past to present are presented below (Figure 5).

Figure 5

Annual Number of Publications and Citation in AI in Educational Research
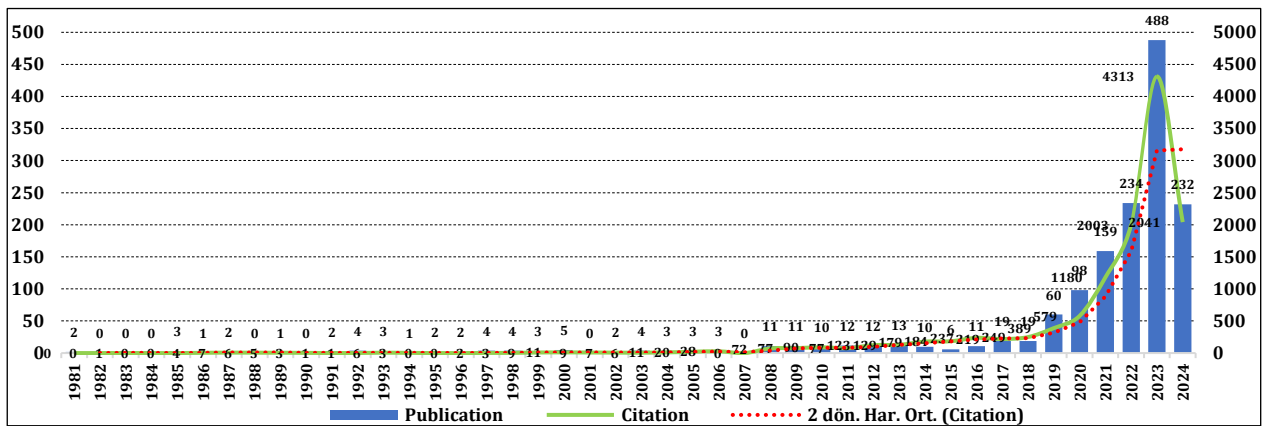
When Figure 5 is examined, it is seen that the number of articles containing AI in education has increased significantly, especially since 2019. There were almost no AI studies until 2019. Therefore, it is clear that there has been an increasing interest in this issue in the last five years. When we look at the number of citations, we see that there have been significant increases since 2019. Although there were very few studies on AI in the field of education between 1981 and 2019, it is noteworthy that there are significant increases in the number of articles and citations as we approach the present day. The table below lists prominent productive authors on AI-related work in education (Table 1).

Table 1

Most Productive Core Authors on AI in Educational Research

Authors	Publications	Publications Fractionalized
Hwang, G. J.	16	5.3
Chui T. K. F.	14	5.5
Chai, C. S.	12	2.6
Gulson, K. N.	9	3.8
Ogata, H.	9	2.1
Su, J. H.	9	3.7
Xie, H. R.	9	2.1
Zou, D.	9	2.1
Tu, Y. F.	8	2.1
Chen, X. L.	7	1.4
Chu, S. K. W.	7	2.1
Dai, Y.	7	2.2
Lin, C. Y.	7	2.2
Mishra, P.	7	2.3
Chen, C. H.	6	1.6

Table 1 shows authors producing on the subject of AI in education. These authors, also called core authors, direct the changes in the field and profoundly influence the studies carried out. Core authors produce approximately half of all publications in a research field and are known in bibliometrics as the Price Law [$M=0.749*(N_{max})^{1/2}$ (M= min. number of articles, N_{max}= number of articles by prolific author)] (Price, 1963; Yeoh et al., 2013). According to the relevant table of core authors, the number of articles published by Hwang, G. J. is the highest (as N_{max}=16). For this reason, the number of articles for which an author should be listed as a core author is calculated as three. The most productive authors are, Chui, (14), Chai, (12), Gulson, (9), Ogata, H. (9), Su, (9), Xie, (9), Zou, (9), Tu, (8), Chen, (7), Chu, (7), Dai, (7), Lin, (7), Mishra, (7) and Chen, (6) respectively. The table below lists the scientific articles that stand out in terms of citation count (Table 2).

Table 2

Most Productive Core Authors on AI in Educational Research

| Paper	Doi	Total Citations	TC per Year	Normalized TC
Chatti, M. A., 2012	10.1504/IJTEL.2012.051815	311	23.92	6.37
Garcia, P., 2007	10.1016/j.compedu.2005.11.017	253	14.06	5.44
Misyak, J. B., 2012	10.1111/j.1467-9922.2010.00626.x	199	15.31	4.08
Cotton, D. R. E., 2024	10.1080/14703297.2023.2190148	190	190.00	145.77
Hwang, G. J., 2003	10.1016/S0360-1315(02)00121-5	183	8.32	2.13
Tlili, A., 2023	10.1186/s40561-023-00237-x	182	91.00	35.21
Kessler, G., 2018	10.1111/flan.12318	139	19.86	5.83
Goralski, M. A., 2020	10.1016/j.ijme.2019.100330	138	27.60	7.47
Chou, C. Y., 2003	10.1016/S0360-1315(02)00130-6	125	5.68	1.46
Chatterjee, S., 2020	10.1007/s10639-020-10159-7	110	22.00	5.95
Farrokhnia, M., 2023	10.1080/14703297.2023.2195846	106	53.00	20.51
Warschauer, M., 2008	10.1080/15544800701771580	105	6.18	4.25
Cooper, G., 2023	10.1007/s10956-023-10039-y	104	52.00	20.12
Smith, R., 2010	10.1177/1046878109334330	99	6.60	3.30
Lim, W. M., 2023	10.1016/j.ijme.2023.100790	98	49.00	18.96

When Table 2 is examined, the most cited article by Chatti et al. (2012) (23.92 citations per year) is the article titled "a reference model for learning analytics". This article is followed by articles written by Garcia et al. (2007) (14.06), Misyak and Christiansen (2012) (15.31), Cotton et al. (2024) (190.00), Hwang (2003) (8.32), Tlili et al. (2023) (91.00), Kessler, (2018) (19.86), Goralski and Tan (2020) (27.60), Chou et al. (2003) (5.68) and Chatterjee and Bhattacharjee (2020) (22.00) respectively. The figure below shows the distribution of authors by country (Figure 6).

Figure 6

Corresponding' Author's Country on AI in Educational Research
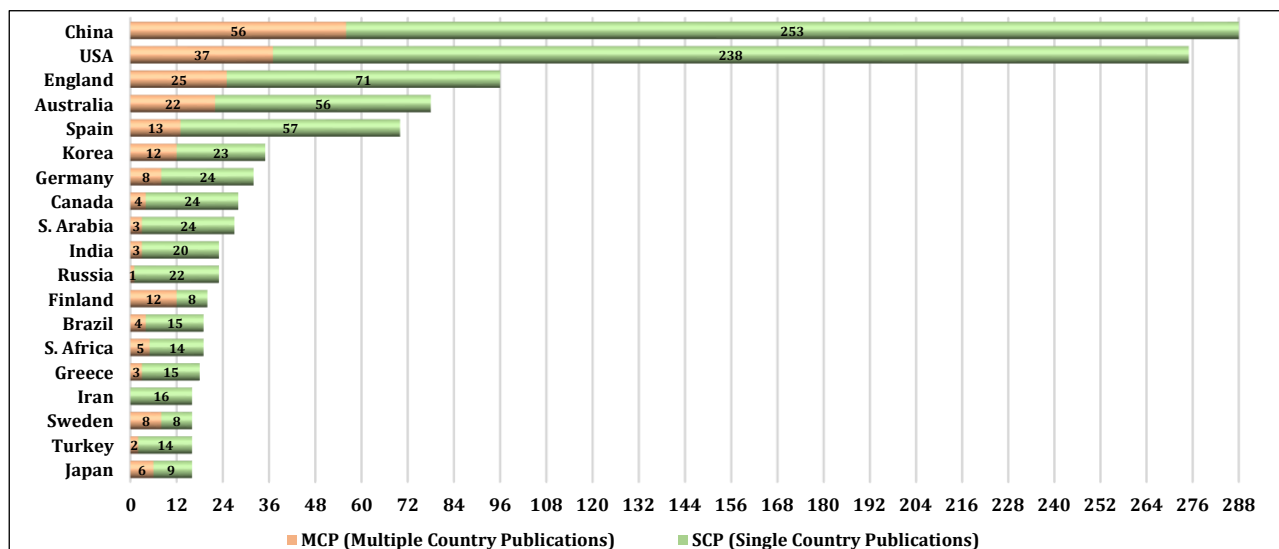
When Figure 6 is examined, the authors who write articles on artificial intelligence in education are primarily from one country. The country with the most connections among authors writing articles on the specified research topic was China (*n*=309). This country is followed by USA (*n*=275), England (*n*=96), Australia (*n*=78), Spain (*n*=70), Korea (*n*=35), Germany (*n*=32), Canada (*n*=28), Saudi Arabia (*n*=27), India (*n*=23), Russia (*n*=23), Finland (*n*=20), Brazil (*n*=19), South Africa (*n*=19), Greece (*n*= 18), Iran (*n*=16), Sweden (*n*=16), Turkey (*n*=16) and Japan (*n*=15) respectively. The figure below shows the authors who took the dominant role in a specific period (Figure 7).

Figure 7

Authors' Production Over Time on AI in Educational Research
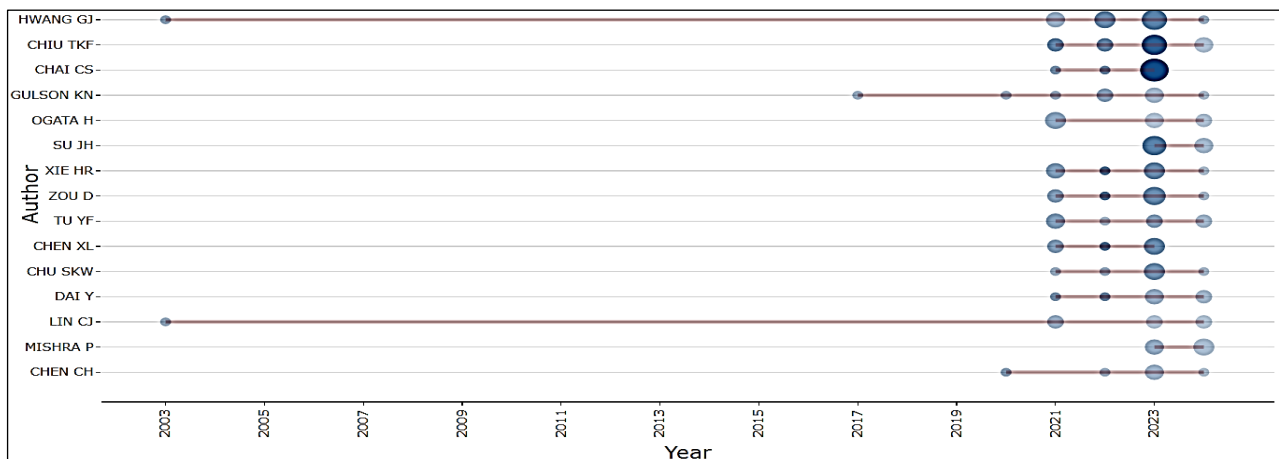
When Figure 7 is examined, the authors named Hwang, G. J. and Lin C. J., among the authors who have played the dominant role from past to present, were quite influential in the relevant field between 2003 and 2024. In the last three years (between 2021 and 2024), Chiu, T. K. F., Chai, C. S., Ogata, H., Su, J. H., Xie, H. R., Zou, D., Tu, Y. F., Chen, X. L., Chu, S. K. W., Dai, Y., authors stand out more. Between 2017 and 2024, authors named Gulson, K. N.; between 2023 and 2024, Mishra, P., took the dominant role. The figure below shows productive institutions on AI in education (Figure 8).

Figure 8

Most Relevant Affiliations Over Time on AI in Educational Research
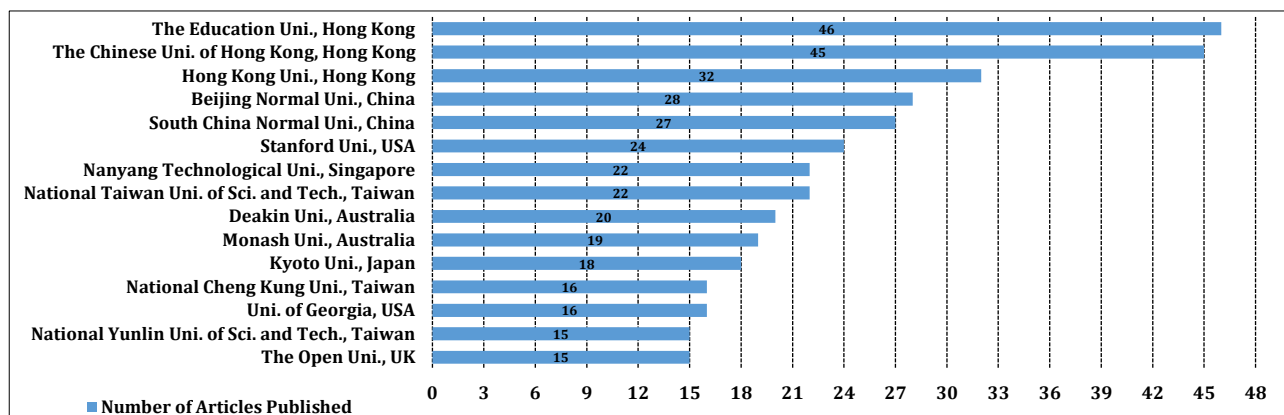
When Figure 8 is examined, The Education (*n*=46) and The Chinese University of Hong Kong (*n*=45) are productive institutions in AI-related work in education. This is followed by institutions; Hong Kong (*n*=32), Beijing Normal (*n*=28), South China Normal (*n*=27), Stanford (*n*=24), Nanyang Technology (*n*=22), National Taiwan (*n*=22), Science and Technology (*n*=22), Deakin (*n*=20), Monash (*n*=19), Kyoto (*n*=18), National Cheng Kung (*n*=16), Georgia (*n*=16), National Yunlin (*n*=15) and The Open (*n*=15) respectively.

3.2. Network analysis on artificial intelligence

3.2.1. Co-Citation networks

Network analysis is an effective technique to reach essential findings in bibliometric research. Visualizations are made by including associations such as authors, countries, and references in the fictional structure of network analyses. In network analyses, co-citation network analyzes generally come to the fore, and the way two scientific articles are cited together is described as co-citation and is

shaped according to nodes and thicknesses in network visualization (Bağış, 2021). The criteria used for the visualization process are stated below the figures (Figure 9).

Figure 9

Co-Cited Network Analysis in the Context of Cited Authors (≥20 articles)
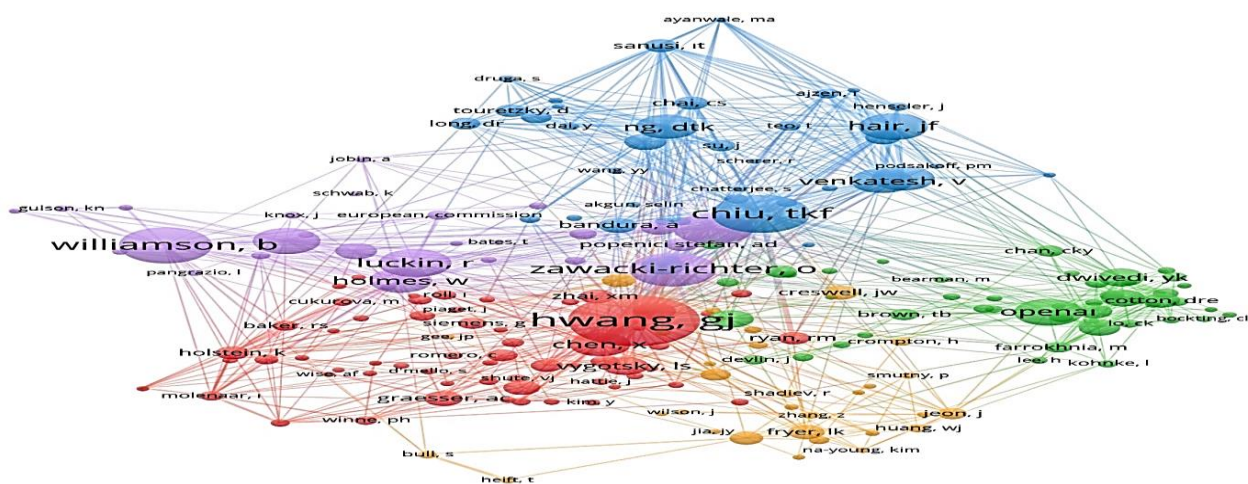
When Figure 9 is examined, according to the co-citation network analysis, network visualization occurred in five different clusters within the context of authors. The nodes formed between clusters and between authors within the cluster show the strength of the connections and indicate the influence of the authors in that cluster according to the node's width (Findlay & van Rensburg, 2018). For example, in the red cluster, Hwang and Chen influence the cluster as dominant authors and write articles on similar topics. On the other hand, in the purple cluster, Zawacki-Richter, Luckin, Williamson, Selwyn; in the green cluster, Openai, Dwivedi, Cotton, Cooper, and Tlili; in the blue cluster, Chiu, Venkatesh Cohen, Ng, and Hair; in the orange cluster, Fryer, Godwin-Jones, Zhang and Jeon authors stand out as the most influential authors of the clusters. These writers influence the cluster they are in and tend to anchor their community by directing both the cluster and other clusters with their study subjects (Mostafa, 2020). Below is a visualization of the co-citation network formed in the context of sources (Figure 10).

Figure 10

Co-Cited Network Analysis in the Context of Cited Sources (≥20 articles)
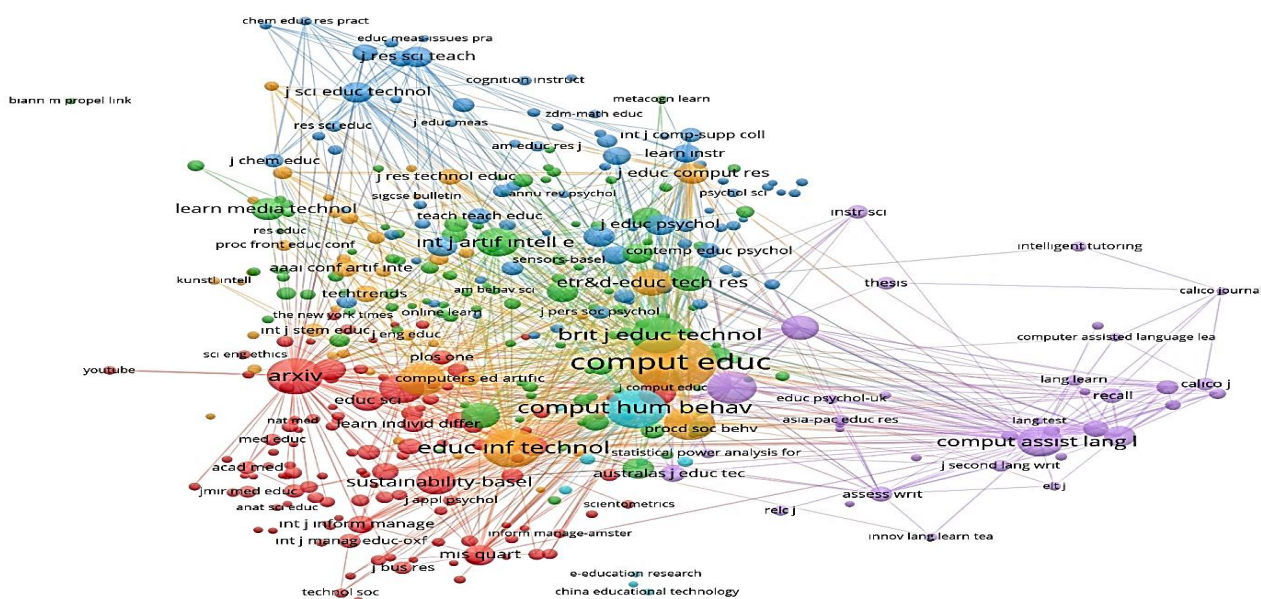
When Figure 10 is examined, according to the co-citation network analysis, network visualization occurs in six different colored clusters in the context of resources. Computer & Education in the orange cluster; Computers in Human Behavior in the light blue cluster; Educational Technology & Society, Computer Assisted Language Learning, Journal of Computer Assisted Learning, and The Language Learning Journal in the purple cluster; International Journal of Educational Technology in Higher Education and Science Education in the red cluster; British Journal of Educational Technology, International Journal of AI in Education in the green cluster; Review of Educational Research, Journal of Research in Science Teaching and Journal of Science Education and Technology in the blue cluster, resources interact intensively in their cluster. In this respect, these resources manage the changes in the field and influence the cluster in which they are located and other clusters.

3.2.2. Collaboration networks

In this part of the research, visualizations of collaborations on AI-related issues in the field of education are included. In this regard, within the scope of co-author analysis, the general view of the collaboration networks between authors, institutions, and countries is presented in visual form. The fictional nature of co-author collaboration means that more than one author contributes to writing a scientific article and naturally becomes a part of the work. In this respect, a detailed view of the social network structure between authors, institutions, and countries is obtained (Bağış, 2021). Below are the co-author network visualizations according to author, institution, and country criteria, respectively.

Figure 11

Co-Authorship Network Analysis in the Context of Authors (≥3 articles)
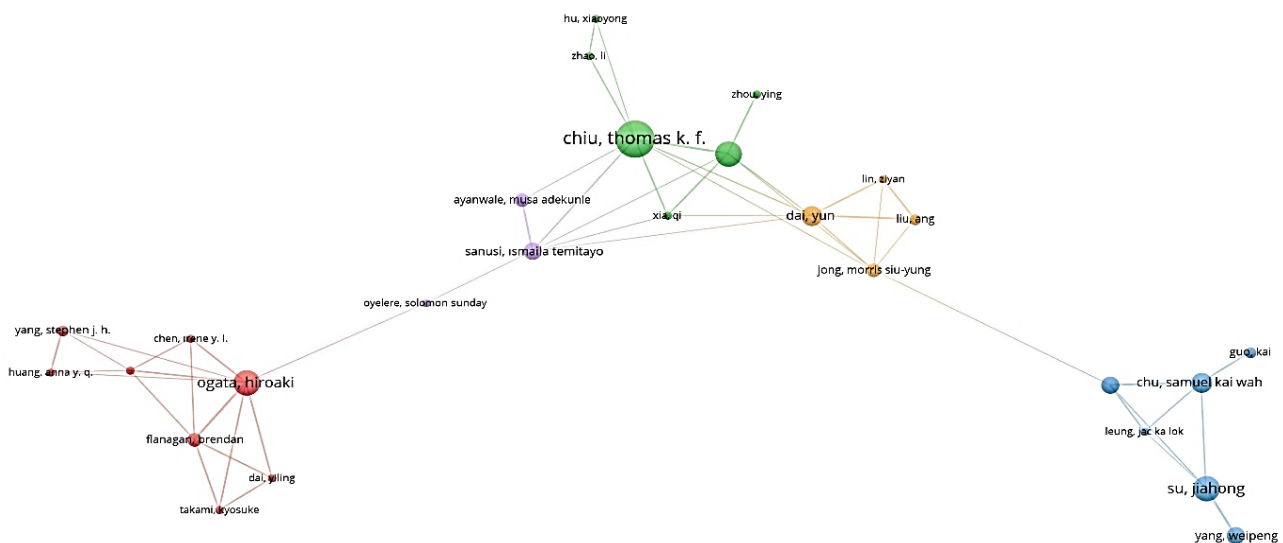
When Figure 11 is examined, it is seen that the collaboration between the authors is collected in five different clusters. However, there is no intense interaction in terms of both the width of the nodes and the connection strength. Accordingly, authors named Chiu dominate the green cluster, Ogata dominates the red cluster, Dai dominates the orange cluster, Su dominates the blue cluster, and Sanusi dominates the purple cluster. These authors have established limited interaction within their cluster. Therefore, there is not much cooperation and interaction between authors on AI in education. Collaborations between authors are generally between the same institution, university, Ministry of Education, and people living in the country, and geographical proximity also deeply affects collaborations. The author or authors who participate in this collaboration and are in a central position are described as "information brokers" and act as information disseminators (Park et al., 2015). Below is a visualization of the co-author network according to institution criteria (Figure 12).

Figure 12

Co-Authorship Network Analysis in the Context of Institutions (≥3 articles)

When Figure 12 is examined, it is seen that the visualization of cooperation between institutions is divided into 17 different clusters. According to network analysis, The Chinese University of Hong Kong, Beijing Normal University, National Taiwan University of Science and Technology, National Cheng Kung University, Hong Kong University, Deakin University, Sydney University, South China Normal University and Monash University appear to be more productive in institutions' collaboration. One of the essential features of these institutions is that they are located in Far Eastern countries. Therefore, more emphasis is placed on AI applications in education, especially in Far Eastern countries. Below is a visualization of the co-author network according to country criteria (Figure 13).

Figure 13

Co-Authorship Network Analysis in the Context of Countries (≥3 articles)

When Figure 13 is examined, it is seen that the visualization of cooperation between countries is divided into 11 different clusters. According to network analysis, China, USA, Taiwan, Australia, England, Spain, India, Canada, Finland, South Africa, and Russia are the more prominent countries in cooperation. China, which is in the purple cluster, has cooperated chiefly with countries that are geographically closer to it, such as Thailand, Japan, Taiwan, and Singapore. The USA has cooperated chiefly with Australia, Canada, Taiwan, Colombia, and the Netherlands in the green cluster. India, which is in the red cluster, mainly cooperates with countries such as Saudi Arabia, Malaysia, Russia, and Bangladesh. According to the results of this analysis, geographical proximity has a significant impact on countries' cooperation. Similarly, the impact of cooperation can be mentioned in the context of developed and developing countries. Below is a visualization of the co-author network according to country criteria (Figure 14).

Figure 14

Authors' World Collaboration Network on AI in Educational Research
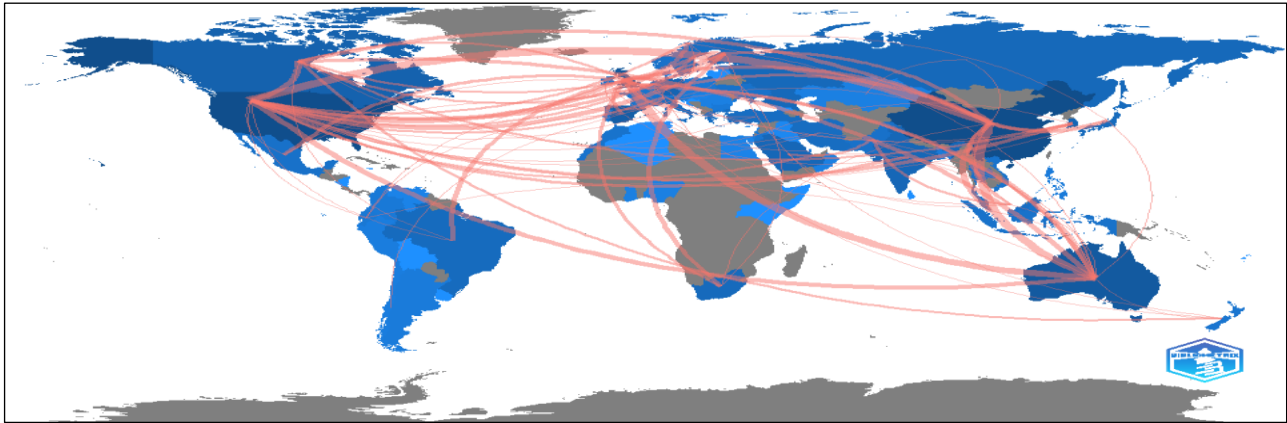

Figure 14 shows intense traffic in the co-network between countries. According to this cooperation network between countries, collaborations between China and the USA (*n*=28), between the UK and Australia (*n*=18), between China and Australia (*n*=16), between USA and Australia (*n*=10), between USA and Canada (*n*=9), between USA and Germany (*n*=8), between Australia and Germany (*n*=7), between China and Canada (*n*=7), between China and Japan (*n*=7), between China and Singapore (*n*=7), between China and UK (*n*=7), between Australia and Canada (*n*=6), and between UK and Germany (*n*=6) are more prominent. According to network relationship traffic, relationships between continents come to the fore. In particular, the intense network between countries in Asia, America, and Australia is noteworthy. According to network relationship traffic, relationships between continents come to the fore. In particular, more intense cooperation between countries in Asia, America, and Australia is noteworthy.

3.2.3. Keywords and co-occurrence network analysis

The keywords of the determined study topic give clues about the article's content and help access the desired documents. Keywords are highly preferred in bibliometric analyses due to their features and the basic framework of the analyses (Chen et al., 2023). In this bibliometric content research based on multifaceted analyses, word cloud analysis was conducted according to the keywords determined by the authors. In the fictional structure of the word cloud, textual data is visualized to obtain a view, and its value in research is determined depending on the predominant use of the word (Liao et al., 2019).

Figure 15

Word Cloud for on AI in Educational Research (≤70, Keywords plus)
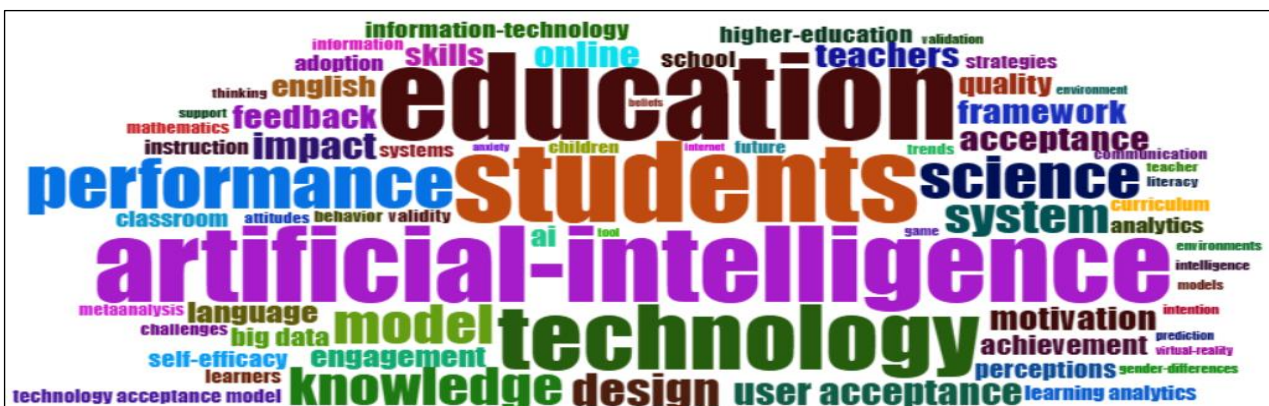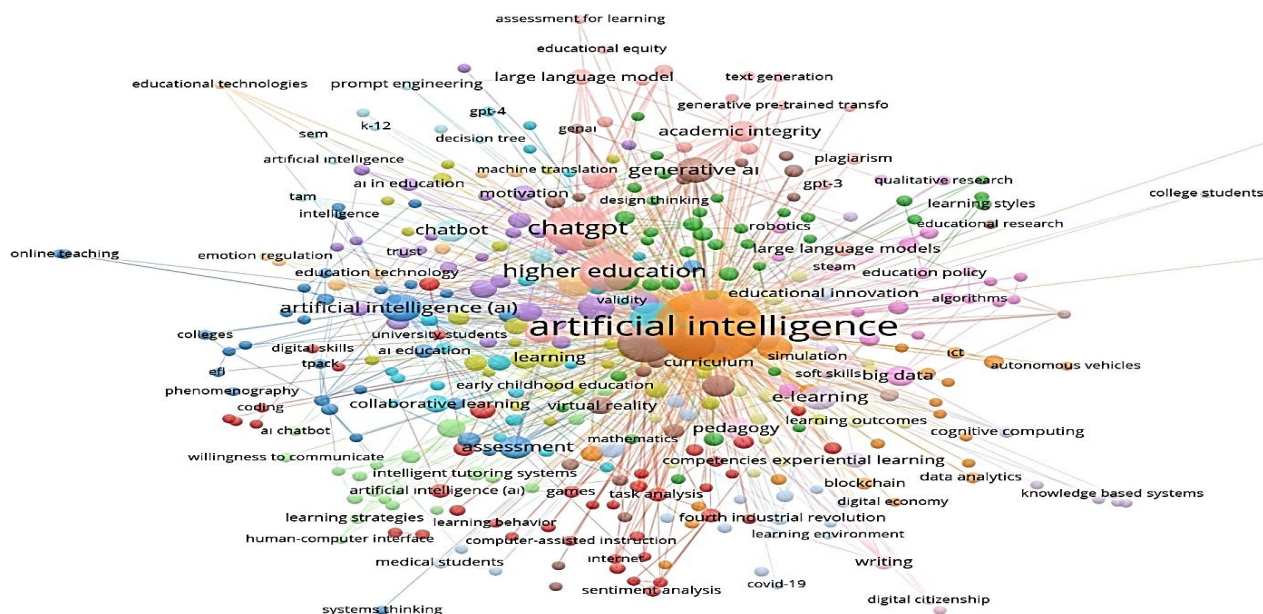

When Figure 15 is examined, the keywords frequently preferred by researchers related to the research topic are; *"education," "students," "artificial intelligence," "technology," "performance," "science," "model,"*

"system," "design," "impact," "knowledge," "online," " user acceptance" and *"teachers."* The number of common words was used to draw the general picture of AI-related research in education to determine in which areas the studies showed more tendencies and how frequently they were used. The visualization obtained from the common word network analysis is presented below (Figure 16).

Figure 16

Co-occurrence Network Analysis in the Context of Author Keywords (≥3)

When Figure 16 is examined, it is seen that many clusters are formed in the keyword network analysis with AI content in the field of education. Especially in the orange cluster, the words *"artificial intelligence"* in the light brown cluster, *"ChatGPT," "higher education,"* in the brown cluster, *"generative artificial intelligence (GAI),"* and in the light blue cluster, *"machine learning"* come to the fore. These words also point to trends regarding the research topic and provide ideas for future researchers. Below is a three-plot (Sankey diagram) analysis (Figure 17). According to the results of the analysis, the authors are located in the middle of the figure, the contents are on the left, and the countries are on the right.

Figure 17

Sankey Plot on AI in Educational Research (keywords- authors-countries)
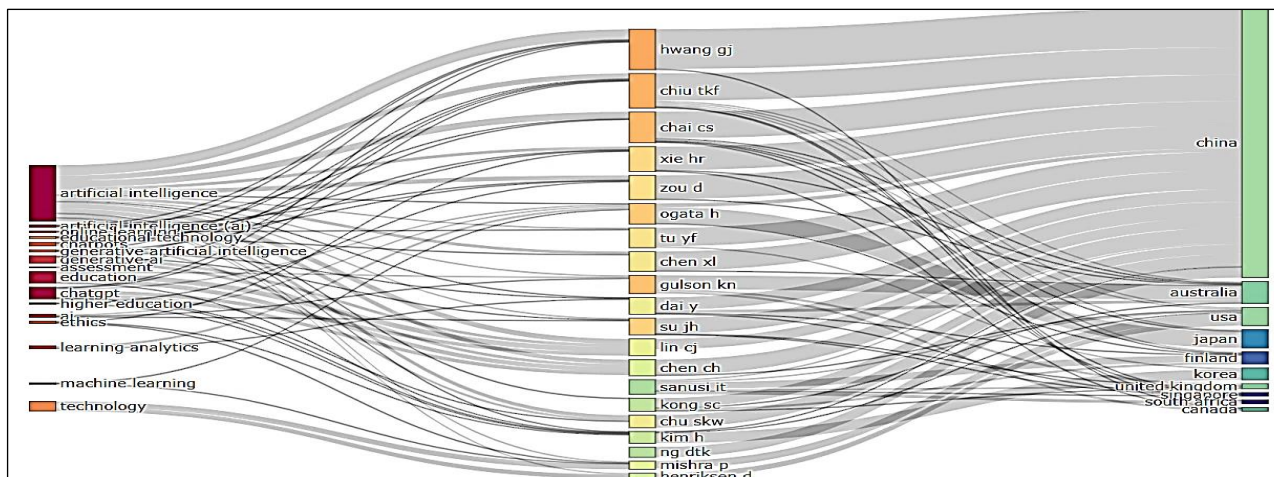
When Figure 17 is examined, the size of the boxes provides us with detailed and holistic information about the content, authors, and countries of the studies related to the research topic. Depending on the size of the boxes, the keyword *"artificial intelligence"* covers many authors and countries in many ways. Likewise, the keywords *"education"* and *"ChatGPT"* also have a widespread effect. When we look at the authors' category, the keywords used by the authors, such as Hwang, G. J., and Chiu T. K. F., significantly affect the field and direct the studies in the field. One of the strengths of studies in the field is that it is demonstrated that the countries that are influential in the research topic (China, Australia, USA, *etc.*) have a say in this field.

3.2.4. Trending topics and thematic evolution

In this section of the findings section, trending (prominent) topics related to the research topic and thematic changes determined depending on the topics are presented holistically. In this heading, the changes in AI-related studies in education from the past to the present have been analyzed over the years. Trend topics and thematic changes show which topics are more prominent in specific periods and hot spots in research (Chen et al., 2023; Mostafa, 2022). Below, the trend topics titles of research with AI content in educational research are presented (Figure 18).

Figure 18

AI in Educational Research Topics Map
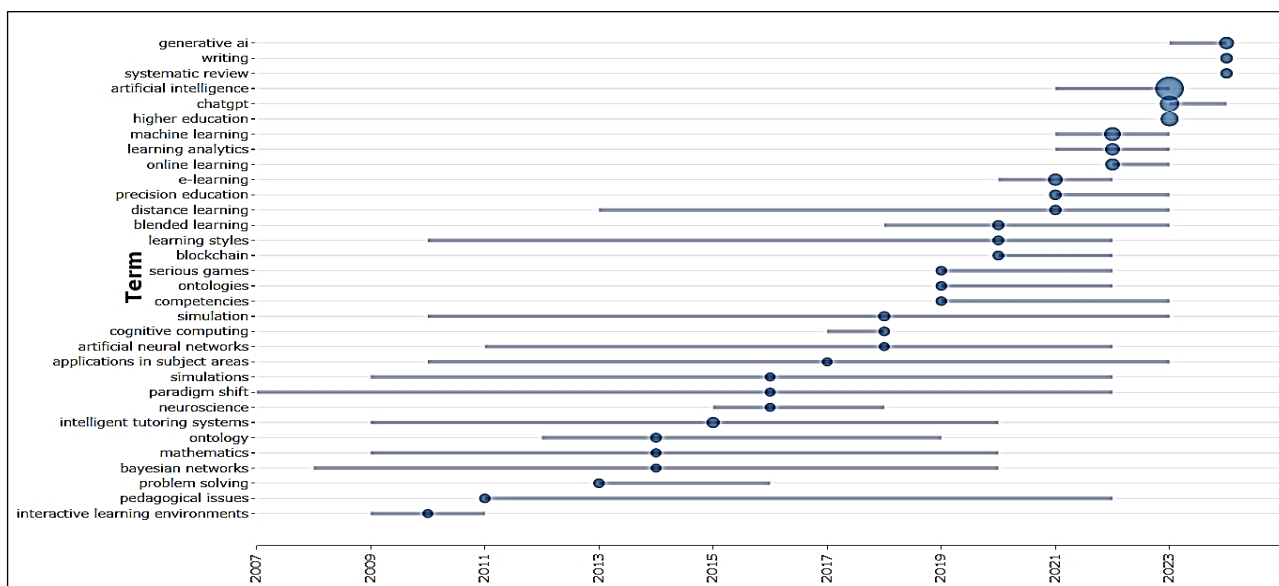
When Figure 18 is examined, it is seen that new generation technological content such as *"AI" (2021-2023), "ChatGPT" (2023-2024), "higher education" (2023-2024), "machine learning" (2021-2023), "GAI" (2023-2024), "learning analytics" (2021-2023), "e-learning" (2021-2023), "online learning" (2020-2022), "intelligent tutoring systems" (2009-2020), "distance learning" (2013-2023), "simulation" (2010-2023), "blended learning" (2018-2023)* and *"blockchain" (2020-2022)* are abundant. As we get closer to today, it is understood that there are trends toward machine learning, blockchain, e-learning, GAI, and learning analytics, as well as AI.

3.3. Conceptual structure and thematic maps

In this part of the research, an attempt was made to determine the general view of the conceptual structure using *Multiple Correspondence Analysis (MCA)* based on the keywords determined by the authors for the research topic. Thanks to this analysis, conceptual mapping was tried to be determined in detail. Therefore, the conceptual structure of AI-containing documents has been visualized since 1981. Below are the MCA findings of scientific articles containing AI in education (Figure 19).

Figure 19

AI in Educational Conceptual Structure Map (MCA Method)
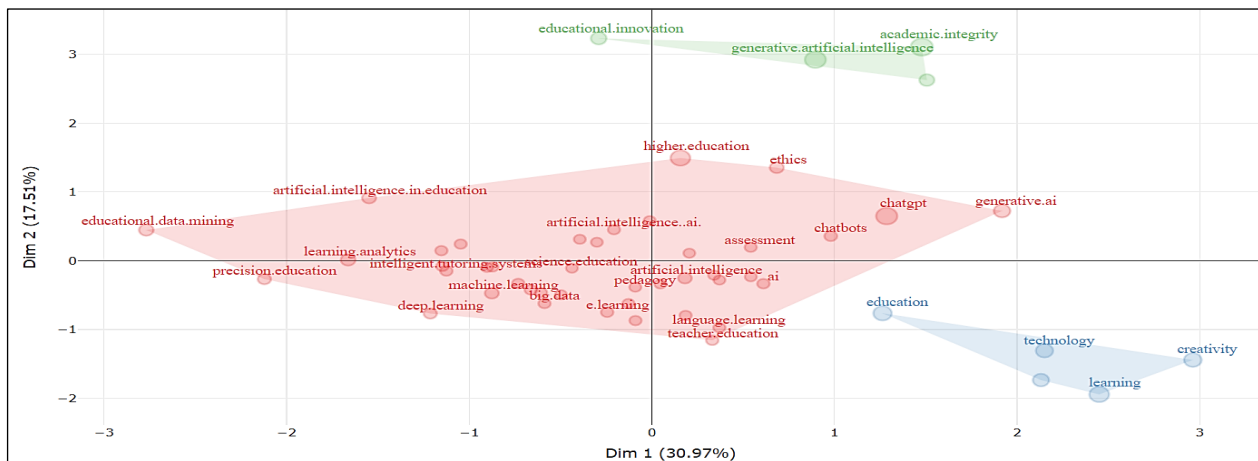
When Figure 19 is examined, according to the MCA result based on the keywords determined by the author(s) related to the research topic, the best dimension reduction was calculated as 48%, depending on three different clusters. According to the figure showing the factor analysis application, the clusters represented contain different colors, and the distance between the points represents common features (Wong et al., 2021). Accordingly, it can be seen that new generation technological concepts such as *"learning analytics," "AI," "ChatGPT," "higher education," "AI in education," "machine learning," "deep learning," "intelligent tutoring systems," "chatbots," "ethics," "GAI"* and *"language learning"* are more intense in the red cluster. Therefore, the effects of technological innovations are felt quite intensely in this cluster. In the blue cluster, more technology and education combinations such as *"education," "technology," "creativity,"* and *"learning"* come to the fore. In the green cluster, innovations in education such as *"educational innovation," "GAI,"* and *"academic integrity"* stand out. Below is the thematic map representing AI-related topics in education (Figure 20).

Figure 20

AI in Educational Thematic Map (Authors Keywords)
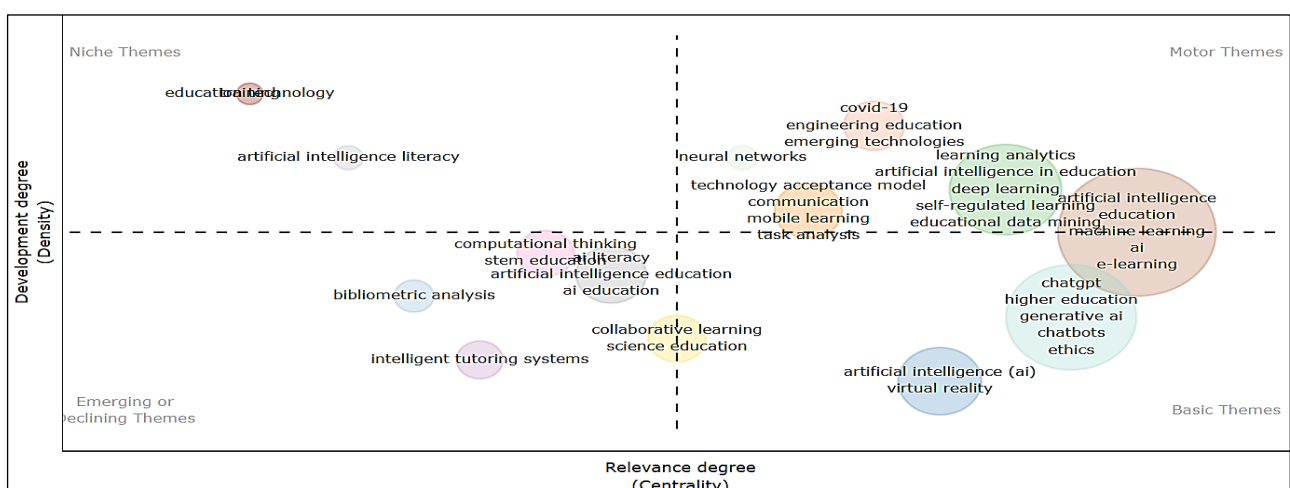
The clustering algorithm Walktrap style was preferred in visualizing the thematic mapping in Figure 20. The map has four quadrants: Niche, Engine, Emerging or Declining, and Basic themes. The bubble size in the image is determined in proportion to the number of keywords preferred by the authors and is interpreted depending on the bubble size (Mostofa, 2022). Motor themes, one of the four quadrants, involve high density and centrality and indicate internal and external development (Cobo et al., 2011).

The representations determined within this theme are the technology acceptance model, communication, mobile learning, task analysis, learning analytics, AI in education, deep learning, AI, education data mining, and self-regulated learning. These representational contents are widely used in the relevant field and cover a wide area of influence. Another quadrant, Niche themes, consists of developed but isolated representations. The representations determined within this theme are education technology and AI literacy. Although the impact areas of these representative contents are limited, they provide valuable information about the content of the study area. Another quadrant, Emerging or Declining themes, consists of low-density and central representations. The representations determined within this theme are as follows: computational thinking, stem education, AI education, bibliometric, and intelligent tutoring systems. Although the impact areas of these representational contents are limited, they represent the center of the study field. Another quadrant is basic themes; it involves low density but high centrality. The representations determined within this theme are as follows: e-learning, generative AI, chatbots, ChatGPT, and virtual reality. This theme also shows the prevalence of the field of study and provides information about the trends of today's AI-related studies.

4. Results, Discussion and Recommendations

Within the scope of this study, a bibliometric analysis of AI-related studies in the field of education from the past to the present was carried out. According to the determined criteria, it has been observed that the increase in the number of articles containing AI in the field of education has been continuous since 1981. It has been determined that the number of scientific articles and citations has gained momentum, especially in the last five years, and similar increases in AI are expected to increase in the coming years. These findings coincide with the results of studies conducted in previous years (Bozkurt et al., 2021; Chen et al., 2022; Peak & Kim, 2021). As AI took place in human history, the development of new technological tools that will profoundly affect social life has accelerated. Therefore, both the increase in the number of scientists and the diversity of AI-based applications indicate that AI will be an essential field of study in the future. Because of the increase in reference numbers, interest in AI is expected to increase. Especially considering that the development of other disciplines depends on education, the effects of AI on education will be inevitable. For this reason, depending on the number of scientific studies and citations in the future, documents related to AI in education will reach significant volumes. Similar studies indicate that the number of studies on AI will increase significantly in the coming years (Bahroun et al., 2023; Forero-Corba & Negre-Bennasar, 2024; Xie et al., 2019).

It has been determined that authors named Hwang, G. J, Chui, T. K. F., Chai, S. S., Gulson, K. N., Ogata, H., Su, J. H., Xie, H. R., and Zou, D. are more prominent in the field of AI in the field of education. These authors direct the developments in the field and lead in determining the content of the study subjects. These authors' common features include receiving many citations, designing qualified studies, and creating reference sources by creating content-rich resources in the field. When the most cited studies were examined, it was determined that the article "A Reference Model for Learning Analytics," published by Chatti et al. (2012), stands out. In this study, the authors focused on learning analytics, which includes many disciplines such as machine learning, AI, information retrieval, statistics, and visualization. Another highly cited study by Garcia et al. (2007) published "Evaluating Bayesian Networks' Precision for Detecting Students' Learning Styles." The authors used the proposed Bayesian model to determine the student's learning style in an AI web-based education system in this study. These studies are seen as pioneering studies in terms of the development and progress of AI. Considering the increasing role of AI in our lives, the need for such studies is expected to increase. In social development, AI-focused technological content is needed more than ever, and valuable steps are being taken to develop productive AI-based applications (OECD, 2019; UNESCO, 2021). For these reasons, AI systems, included in people's daily lives in all segments of society, will continue to become an indispensable part of life.

Another research finding was obtained from the distribution of responsible authors by country. Corresponding authors are mostly affiliated with China, the USA, the UK, Australia, Spain, and Korea. Therefore, there is intense interest in AI in education studies in these countries. On the other hand, many different countries are interested in this research topic. This finding parallels studies in the literature (Hinojo-Lucena et al., 2019). The majority of writers stand out as writers from a single country. These authors generally participate in joint studies based on geographical proximity and produce a limited number of studies in the context of multi-country authors. Especially in the period close to the present day, authors named Hwang, G., Chiu, T. K. F., and Chai, C. S. have taken the dominant role. These authors are also seen as core authors and profoundly influence change in the field. One of the remarkable findings of the research is that institutions in Hong Kong have come to the fore in AI-related studies in the field of education. Many influential authors are affiliated with these institutions (The Education University, The Chinese University, Hong Kong University) and collaborate intensively on published quality studies. Network analysis reveals the invisible collaborative network of countries and institutions participating in research through scientific collaboration visualization and helps intuitively display potential scientific collaborators (Chen et al., 2022). In the findings obtained from network analyses, authors named Hwang, Chen, Zawacki-Richter, Luckin, Williamson, Chiu, Openai, Dwivedi, Cotton, and Fryer determined that the connection clusters in the network visualization were thick and frequent. Therefore, these authors unite on similar subject contents and stand out more on the topics in their fields of study. The subject areas of these authors include content that will guide the field, such as AI application examples, the theoretical framework in AI-based learning, AI-based learning outcomes, AI-designed learning experiences, and the effectiveness of AI in application areas. In terms of resources, there is heavy collaboration traffic between Computer & Education, Computer Assisted Language Learning, Computers in Human Behavior, Educational Technology & Society, Journal of Computer Assisted Learning, British Journal of Educational Technology, International Journal of Educational Technology in Higher Education and Science Education, Journal of Research in Science Teaching, Journal of Science Education and Technology, Review of Educational Research and International Journal of AI in Education resources. These sources are prestigious and have gained a respected place around the world. Most of these sources, which contain qualified studies, have SSCI, SSCI-Expanded, and ESCI indexes and are scanned in the WoS database (WoSG, 2024). Another result of the research findings is that countries from many continents cooperate in the cooperation network. Especially China and the USA stand out in this field. Although this finding is expected, it is known that the Ministry of Education of China has intensive studies in AI (Cui et al., 2018; Huang et al., 2021b; Shi et al., 2024). Similarly, many publishing organizations are based in the USA and host leading studies in this field (OECD, 2019).

According to the findings obtained from the trending topic and thematic change heading of the research, it has been determined that as we get closer to the present day, ChatGPT, generative AI, machine learning, deep learning, blockchain, AI in ethics, language learning chatbots, intelligent tutoring systems and learning analytics topics become prominent. These issues are among the current issues in the field. AI has permeated many areas of our lives with the development of technology. It has become an essential part of our social lives. Therefore, the presence of AI in educational environments will be felt more and more day by day (Chen et al., 2020; Huang et al., 2021a; Hwang et al., 2020; Lin et al., 2021; Russell & Norvig, 2021). AI technology has emerged as a revolutionary force that has profoundly transformed various aspects of human life (Bahroun et al., 2023). The research findings show that AI is addressed in education in many ways. Developments resulting in the development and use of AI in education provide innovative opportunities for researchers to benefit from AI. Therefore, when the effects of AI in different subject disciplines are evaluated together, the subject diversity in the field will expand further. Such bibliometric studies provide valuable clues in determining trends, contents, productions, and collaborations in the field. Therefore, periodically reviewing publications related to the research topic is necessary.

4.1. Limitations and agenda and implications for future research

This study aims to examine the current landscape of AI in education, predict its future effects, and explore its future direction. Conducting longitudinal studies emphasizing curriculum designs to ensure the full integration of AI into educational environments can provide qualified data for future research. Its effects over long periods can be especially revealed with the help of longitudinal studies. The keywords presented by the researchers revealed a clear pattern regarding AI in education, which has yet to escape the field of traditional education. Therefore, in-depth research on the direct effects of AI algorithms and tools on education can be further encouraged. The emergence of AI technologies in education has begun to manifest itself intensively in areas such as medicine and health, engineering, economy, transportation, agriculture, and logistics. Therefore, its relationship with the field of education can be investigated by conducting similar studies in different disciplines. In particular, the impact of AI-based tools in learning environments can be examined through experimental research, and their role in shaping the future of education can be revealed in depth. The findings obtained from the research offer valuable opportunities for the integration of AI into education and the diversification of educational content. Considering the findings of studies involving AI, researchers can conduct studies on the effects of educational materials with more dynamic content by focusing on individual learning. More emphasis can be placed on subject areas that will shape the lifelong learning experience of AI applications. Additionally, learning designs can be created to increase AI awareness. Topics such as ChatGPT, machine learning, deep learning, and blockchain can be researched by considering many factors, such as students' achievements, interests, performances, perceptions, and affective changes. The effects of collaborative learning applications that center on the collaboration of humans and AI can be determined to achieve educational outcomes, which is another gap in the field. In addition, their effects on the performance of students with specific learning difficulties can also be examined. This study, which deals with the content of AI in education holistically, also has certain limitations. The most important limitations of the study include the fact that only the WoS database was used, Education Educational Research was chosen as the WoS category, only studies written in English were included in the research, and only scientific articles were used as the data set.

References

Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics,* 11(4), 959-975. https://doi.org/10.1016/j.joi.2017.08.007

Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology,* 8(1), 19-32. https://doi.org/10.1080/1364557032000119616

Bağış, M. (2021). Main analysis techniques used in bibliometric research. In O. Öztürk & G. Gürler (Eds.) *Bibliometric analysis as a literature review tool* (pp. 97-123). Nobel Academic Publishing.

Bahroun, Z., Anane, C., Ahmed, V., & Zacca, A. (2023). Transforming education: a comprehensive review of generative artificial intelligence in educational settings through bibliometric and content analysis. *Sustainability,* 15, 1-40. https://doi.org/10.3390/su151712983

Bayne, S. (2015). Teacherbot: Interventions in automated teaching. *Teaching in Higher Education,* 20(4), 455-467. https://doi.org/10.1080/13562517.2015.1020783

Bozkurt, A., Karadeniz, A., Baneres, D., Guerrero-Roldán, A. E., & Rodríguez, M. E. (2021). Artificial intelligence and reflections from educational landscape: A review of AI studies in half a century. *Sustainability,* 12(2), 1-16. https://doi.org/10.3390/su13020800

Chaparro, N., & Rojas-Galeano, S. (2021). *Revealing the research landscape of master's degrees via bibliometric analyses. Library Philosophy and Practice,* arXiv:2103.09431. https://doi.org/10.48550/arXiv.2103.09431

Chatti, M. A., Dyckhoff, A. L., Schoeder, U., & Thüs, H. (2012). A reference model for learning analytics. International *Journal of Technology Enhanced Learning,* 4(5/6), 318-331. https://doi.org/10.1504/IJTEL.2012.051815

Chen, H. E., Sun, D., Hsu, T. C., Yang, Y., & Sun, J. (2023). Visualising trends in computational thinking research from 2012 to 2021: A bibliometric analysis. *Thinking Skills and Creativity,* 47, 1-18. https://doi.org/10.1016/j.tsc.2022.101224

Chen, X., Zou, D., Xie, H., Cheng, G., & Liu, C. (2022). Two decades of artificial intelligence in education: Contributors, collaborations, research topics, challenges, and future directions. *Educational Technology & Society,* 25(1), 28-47. https://doi.org/10.30191/ETS.202201_25(1).0003

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

Chen, X., Yu, G., Cheng, G., & Hao, T. (2019). Research topics, author profiles, and collaboration networks in the top-ranked journal on educational technology over the past 40 years: A bibliometric analysis. *Journal of Computers in Education,* 6(4), 563-585. https://doi.org/10.1007/s40692-019-00149-1

Chiu, T. K. F., & Chai, C. S. (2020). Sustainable curriculum planning for artificial intelligence education: A self-determination theory perspective. *Sustainability,* 12(14), 1-18. https://doi.org/10.3390/su12145568

Cobo, M., Lopez-Herrera, A., Herrera-Viedma, E., & Herrera, F. (2011). An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the fuzzy sets theory field. *Journal of Informetrics,* 5(1), 146-166. https://doi.org/10.1016/j.joi.2010.10.002

Cui, W., Xue, Z., & Thai, K. P. (2018, November). *Performance comparison of an AI-based adaptive learning system in China.* In 2018 Chinese automation congress (CAC) (pp. 3170-3175). IEEE.

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research,* 133, 285-296. https://doi.org/10.1016/j.jbusres.2021.04.070

Fang, C., Zhang, J., & Qiu, W. (2017). Online classified advertising: A review and bibliometric analysis. *Scientometrics,* 113(3), 1481-1511. https://doi.org/10.1007/s11192-017-2524-6

Findlay, K., & van Rensburg, O. (2018). Using interaction networks to map communities on Twitter. *International Journal of Market Research,* 60(2), 169-189. https://doi.org/10.1177/1470785317753025

Forero-Corba, W., & Negre-Bennasar, F. (2024). Techniques and applications of machine learning and artificial intelligence in education: A systematic review. *RIED-Revista Iberoamericana de Educación a Distancia,* 27(1), 209-253. https://doi.org/10.5944/ried.27.1.37491

Garcia, P., Amandi, A., Schiaffino, S., & Campo, M. (2007). Evaluating Bayesian networks' precision for detecting students' learning styles. *Computers & Education,* 49(3), 794-808. https://doi.org/10.1016/j.compedu.2005.11.017

Gonzales-Valiente, C. (2019). Redes de citación de revistas iberoamericanas de bibliotecología y ciencia de la información en Scopus [Citation networks of Ibero-American library and information science journals in Scopus]. *Bibliotecas Anales de Investigación,* 15, 83-98. Retrieved from http://agora.edu.es/servlet/articulo?codigo=7871010

Grzybowska, K., & Awasthi, A. (2020). Literature review on sustainable logistics and sustainable production for industry 4.0. In K., Grzybowska, A., Awasthi, & R., Sawhney (Eds.), *Sustainable logistics and production in industry 4.0 new opportunities and challenges* (pp. 1-19). Springer Publishing.

Harmon, J., Pitt, V., Summons, P., & Inder, K. J. (2021). Use of artificial intelligence and virtual reality within clinical simulation for nursing pain education: A scoping review. *Nurse Education Today,* 97, 1-9. https://doi.org/10.1016/j.nedt.2020.104700

Hinojo-Lucena, F. J., Aznar-Díaz, I., Cáceres-Reche, M. P., & Romero-Rodríguez, J. M. (2019). Artificial intelligence in higher education: A bibliometric study on its impact in the scientific literature. *Education Sciences,* 9(1), 1-9. https://doi.org/10.3390/educsci9010051

Huang, X. (2021). Aims for cultivating students' key competencies based on artificial intelligence education in China. *Education and Information Technologies,* 26, 5127-5147. https://doi.org/10.1007/s10639-021-10530-2

Huang, X., & Qiao, C. (2024). Enhancing computational thinking skills through artificial intelligence education at a STEAM high school. *Science & Education,* 33(2), 383-403. https://doi.org/10.1007/s11191-022-00392-6

Huang, J., Shen, G., & Ren, X. (2021a). Connotation analysis and paradigm shift of teaching design under artificial intelligence technology. *International Journal of Emerging Technologies in Learning (iJET),* 16(5), 73-86. https://doi.org/10.3991/ijet.v16i05.20287

Huang, J., Saleh, S., & Liu, Y. (2021b). A review on artificial intelligence in education. *Academic Journal of Interdisciplinary Studies,* 10(3), 206-217. https://doi.org/10.36941/ajis-2021-0077

Hwang, G. J., & Tu, Y. F. (2021). Roles and research trends of artificial intelligence in mathematics education: A bibliometric mapping analysis and systematic review. *Mathematics,* 9(6), 1-19. https://doi.org/10.3390/math9060584

Hwang, G. J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of artificial intelligence in education. Computers and Education: *Artificial Intelligence,* 1, 1-5. https://doi.org/10.1016/j.caeai.2020.100001

Jamali, S. M., Ebrahim, N. A., & Jamali, F. (2022). The role of STEM education in improving the quality of education: A bibliometric study. *International Journal of Technology and Design Education, Springer Verlag,* 32(3), 1-22. https://doi.org/10.1007/s10798-022-09762-1

Kohonen, T. (2001). *Self-organizing maps* (3rd ed.). Springer-Verlag. http://dx.doi.org/10.1007/978-3-642-56927-2

Law, J., Bauin, S., Courtial, J., & Wittaker, J. (1988). Policy and the mapping of scientific change: A co-word analysis of research into environmental acidification. *Scientometrics,* 14(3-4), 251-264. https://doi.org/10.1007/BF02020078

Lee, H. S., & Lee, J. (2021). Applying artificial intelligence in physical education and future perspectives. *Sustainability,* 13(1), 1-16. https://doi.org/10.3390/su13010351

Lemaignan, S., Warnier, M., Sisbot, E. A., Clodic, A., & Alami, R. (2017). Artificial cognition for social human–robot interaction: An implementation. *Artificial Intelligence,* 247, 45–69. https://doi.org/10.1016/j.artint.2016.07.002

Li, Z., & Wang, H. (2021). The effectiveness of physical education teaching in college based on Artificial intelligence methods. *Journal of Intelligent & Fuzzy Systems,* 40(2), 3301-3311. https://doi.org/10.3233/JIFS-189370

Liao, H., Tang, M., Li, Z., & Lev, B. (2019). Bibliometric analysis for highly cited papers in operations research and management science from 2008 to 2017 based on essential science indicators. *Omega,* 88, 223-236. https://doi.org/10.1016/j.omega.2018.11.005

Lin, Y. S., Chen, S. Y., Tsai, C. W., & Lai, Y. H. (2021). Exploring computational thinking skills training through augmented reality and AIoT learning. *Frontiers in Psychology,* 12, 1-9. https://doi.org/10.3389/fpsyg.2021.640115

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics,* 5(4), 115-133. https://doi.org/10.1007/BF02478259

Mostafa, M. M. (2022). Three decades of interactive learning environments: A retrospective bibliometric network analysis. *Interactive Learning Environments,* 31(10), 6968-6987. https://doi.org/10.1080/10494820.2022.2057548

Mostafa, M. M. (2020). A knowledge domain visualization review of thirty years of halal food research: Themes, trends and knowledge structure. *Trends in Food Science & Technology,* 99, 660-677. https://doi.org/10.1016/j.tifs.2020.03.022

Organisation for Economic Co-operation and Development (OECD) (2019). *Artificial intelligence in society* (pp. 47-80). OECD Publishing. https://doi.org/10.1787/eedfee77-en

Paek, S., & Kim, N. (2021). Analysis of worldwide research trends on the impact of artificial intelligence in education. *Sustainability,* 13(14), 1-20. https://doi.org/10.3390/su13147941

Park, S., Lim, Y., & Park, H. (2015). Comparing Twitter and YouTube networks in information diffusion: The case of the "occupy wall street" movement. *Technological Forecasting and Social Change,* 95(6), 208-217. https://doi.org/10.1016/j.techfore.2015.02.003

Popenici, S. A. D., & Kerr, S. (2017). Exploring the impact ofartificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning,* 12(1), 1-13. https://doi.org/10.1186/s41039-017-0062-8

Price, D. J. (1963). *Little science, big science.* Columbia University Press.

Pritchard, A. (1969). Statistical bibliography or bibliometrics. *Journal of Documentation,* 25(4), 348-349. Retrieved from https://cir.nii.ac.jp/crid/1570009750342049664

Pretorius, L. (2023). Fostering AI literacy: A teaching practice reflection. *Journal of Academic Language & Learning,* 17(1), 1-8. Retrieved from

https://research.monash.edu/en/publications/fostering-ai-literacy-a-teaching-practice-reflection

Russell, S. J., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson Publishing.

Shan, S., & Liu, Y. (2021). Blended teaching design of college students' mental health education course based on artificial intelligence flipped class. *Mathematical Problems in Engineering, 2021,* 1-10. https://doi.org/10.1155/2021/6679732

Shi, S. J., Li, J. W., & Zhang, R. (2024). A study on the impact of generative artificial intelligence supported situational interactive teaching on students' 'flow' experience and learning effectiveness-a case study of legal education in China. *Asia Pasific Journal of Education,* 44(1), 112-138. https://doi.org/10.1080/02188791.2024.2305161

Tsai, S. C., Chen, C. H., Shiao, Y. T., Ciou, J. S., & Wu, T. N. (2020). Precision education with statistical learning and deep learning: A case study in Taiwan. *International Journal of Educational Technology in Higher Education,* 17(1), 1-13. https://doi.org/10.1186/s41239-020-00186-2

Tobler, S. (2024). Smart grading: A generative AI-based tool for knowledge-grounded answer evaluation in educational assessments. *MethodsX,* 12, 1-6. https://doi.org/10.1016/j.mex.2023.102531

United Nations Educational, Scientific and Cultural Organization (UNESCO) (2021). *Intergovernmental meeting of experts (category ll) related to a draft recommendation on the ethics of artificial intelligence.* Retrieved from https://unesdoc.unesco.org/ark:/48223/pf0000376712/PDF/376-712-eng.pdf.multi

Van Eck N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics,* 84(2), 523-538. https://doi.org/10.1007/s11192-009-0146-3

Verbeek, A., Debackere, K., Luwel, M., & Zimmermann, E. (2002). Measuring progress and evolution in science and technology-I: The multiple uses of bibliometric indicators. *International Journal of Management Reviews,* 4(2), 179-211. https://doi.org/10.1111/1468-2370.00083

Wang, S., Sun, Z., & Chen, Y. (2023). Effects of higher education institutes' artificial intelligence capability on students' self-efficacy, creativity and learning performance. *Education and Information Technologies,* 28, 4919-4939. https://doi.org/10.1007/s10639-022-11338-4

Web of Science Group (WoSG) (2024). *Web of Science Core Collection.* Retrieved from https://clarivate.com/

Wetzstein, A., Feisel, E., Hartmann, E., & Benton, W. (2019). Uncovering the supplier selection knowledge structure: A systematic citation network analysis from 1991 to 2017. *Journal of Purchasing and Supply Management,* 25(4), 1-16. https://doi.org/10.1016/j.pursup.2018.10.002

Wong, W., Mittas, N., Arvanitou, E., & Li, Y. (2021). A bibliometric assessment of software engineering themes, scholars and institutions (2013-2020). *Journal of Systems and Software,* 180, 1-10. https://doi.org/10.1016/j.jss.2021.111029

Xie, H., Chu, H. C., Hwang, G. J., & Wang, C. C. (2019). Trends and development in technology-enhanced adaptive/ personalized learning: A systematic review of journal publications from 2007 to 2017. *Computers & Education,* 140, 1-16. https://doi.org/10.1016/j.compedu.2019.103599

Yang, J., & Zhang, B. (2019). Artificial intelligence in intelligent tutoring robots: A systematic review and design guidelines. *Applied Sciences,* 9(10), 1-18. https://doi.org/10.3390/app9102078

Yeoh, W., Talburt, J., & Zhou, Y. (2013). *Information quality and governance for business intelligence.* IGI Global Publishing.

Yuan, B. Z., Bie, Z. L., & Sun, J. (2021). Bibliometric analysis of global research on muskmelon (Cucumis melo L.) based on web of science. *Hort Science,* 56(8), 867-874. https://doi.org/10.21273/HORTSCI15827-21

Zhao, S., Shen, Y., & Qi, Z. (2023). Research on chatgpt-driven advanced mathematics course. *Academic Journal of Mathematical Sciences,* 4(5), 42-47. https://doi.org/10.25236/AJMS.2023.040506

Zhao, D., & Strotmann, A. (2015). Analysis and visualization of citation networks. *Synthesis Lectures on Information Concepts, Retrieval, and Services,* 7(1), 1-207. https://doi.org/10.1007/978-3-031-02291-3

Article Information Form

Author Notes: The author would like to express his sincere thanks to the editor and the anonymous reviewers for his helpful comments and suggestions.

Author Contributions: The article has a single author. The author has read and approved the final manuscript.

Conflict of Interest Disclosure: No potential conflict of interest was declared by the author.

Copyright Statement: Author owns the copyright of his work published in the journal and his work is published under the CC BY-NC 4.0 license.

Supporting/Supporting Organizations: No grants were received from any public, private or non-profit organizations for this research.

Ethical Approval: It is declared that during the preparation process of this study, scientific and ethical principles were followed and all the studies benefited from are stated in the bibliography. No human subjects were used in study. It includes bibliometric analysis of data obtained from the literature. For these reasons, ethics committee approval is not required.

Plagiarism Statement: This article has been scanned by iThenticate.

Comparison of Item Difficulty Analyses of Exams Used in Teaching Turkish as a Foreign Language with Instructors' Perceptions of Item Difficulty

Funda Keskin[1*] [ID]
Seçil Alaca[2] [ID]

[1] Sakarya University, Institute of Educational Sciences, Sakarya, Türkiye, funda.keskin@istanbul.edu.tr

[2] Istanbul University, Language Center, İstanbul, Türkiye, secilalaca@gmail.com

*Corresponding author

Abstract: In numerous studies focusing on assessment and evaluation of teaching Turkish as a foreign language, researchers have frequently identified issues related to the standardization and low validity and reliability of exams. Addressing these issues and investigating the underlying causes is paramount. Given the development of assessment tools by Turkish language teaching centers are typically the responsibility of instructors, it is essential to understand their perspectives regarding these tools. This study aimed to evaluate the perceptions of instructors concerning item difficulty in the context of teaching Turkish as a foreign language. Initially, item analyses were conducted on reading tests included in assessment tools designed by a Turkish language teaching center for B1, B2, and C1 proficiency levels. Instructors from various Turkish language teaching centers were asked to evaluate item difficulty through a prepared questionnaire. Data regarding instructors educational backgrounds, experiences, and involvement in exam creation were collected. Various analytical methods were employed to examine and interpret the obtained data. Item analysis results of examined tests were compared with instructors' perceptions of difficulty using fit analysis. Accuracy of instructors' item difficulty estimates was calculated for each instructor using Error Matrix, and success rates determined. To identify the effects of instructors' characteristics on item difficulty estimation, t-test and ANOVA analyses were performed. These analyses results were interpreted alongside item analyses, and recommendations provided to enhance the assessment and evaluation literacy of instructors teaching Turkish as a foreign language.

Keywords: Turkish as a Foreign Language, Item Analysis, Item Difficulty Perception, Assessment in Teaching Turkish as a Foreign Language

1. Introduction

In language tests, item writers can be an extremely important aspect which directly affects test validity, such as their impact on test and test specification development. No matter how high the theoretical validity of a developed test, item writers competence can directly affect test validity. Therefore, the performance of item writers and their accurate guidance is an important stage of test development. To improve item writer performance, it is necessary to first identify the situation. Importantly, the appropriate guidance needed by question writers is enhanced by the performance of item writers. One of the best ways to determine item writer performance is through actual item writing and piloting, even though this is not always practical. For this reason, several studies have relied on expert judgement of item difficulty to determine performance indicators (Fergadiotis et all., 2019; Hambleton & Jirka, 2006; Sydorenko, 2011; Wise et all., 2009).

In the test development process, having knowledge regarding item difficulty is crucial. For a particular group of test participants, the difficulty of items can be determined fairly accurately following pilot testing. However, it is necessary to understand what makes items more or less difficult while being developed and prior to being piloted. A common practice in test development is to provide item writers with item-level descriptors, which are usually grounded in previous research related to predictors of item difficulty. For example, previous research has identified various factors that affect item difficulty including negation in the stems (e.g., Hambleton & Jirka, 2006), topic familiarity (e.g., Freedle & Kostin, 1999), or lexical knowledge (e.g., Rupp, Garcia, & Jamieson, 2001). One of the purposes of such research is to increase item-writing process efficiency by providing specific guidelines and item-level descriptors to item writers (Kostin, 2004). Additionally, further studies have expanded on these factors. For

instance, Bachman (2002) highlighted even experienced item writers often struggle with accurately predicting item difficulty, emphasizing the need for extensive training and detailed guidelines. Similarly, Alderson (1993) found judges are better at predicting the difficulty of reading comprehension items compared to other types, such as cloze tests, indicating variability in difficulty prediction across different item formats. Moreover, Bejar (1983) demonstrated the accuracy of item difficulty estimates can be significantly improved through use of anchor-based methods and training, though these improvements are not always sufficient to entirely replace empirical pretesting. Furthermore, Shohamy (1984) indicated the type of response format (i.e., multiple-choice vs. true-false) significantly impacts item difficulty, with multiple-choice items generally being more challenging due to lower probability of guessing correctly. This body of research collectively aimed to refine the item development process by providing item writers with robust, empirically validated tools, and guidelines which enhance the predictive accuracy of item difficulty. By integrating these findings, test developers can create more reliable and valid assessments to more accurately measure intended skills and knowledge.

Alderson, Clapman, and Wall (1995) argued that no matter how well a test is designed, it can be quite challenging to determine whether the items are appropriate without first piloting them on learners. Even experienced teachers and test experts often disagree on what a specific item measures or how difficult it may be. Therefore, piloting is essential to assess test validity and reliability. Boylu (2019) notes that while item analysis might seem challenging, instructors can easily conduct difficulty and discrimination analyses. However, many educators feel they lack proficiency in areas such as "calculating item discrimination", "determining which items to include based on discrimination index scores", and "calculating test reliability and validity" (Altıntaş, 2022; Boylu, 2019).

Exams created by Turkish Language Teaching Centres (TLTC) often lack validity and reliability due to instructors' limited knowledge regarding item writing and language assessment (Gedik, 2017; Işıkoğlu, 2015). This leads to inaccurate assessments which can negatively affect students (Kutlu et al., 2010). Although the inadequacy of assessment tools and instructors' difficulties in applying language assessment principles have been identified, studies focusing on the root cause of these issues remain insufficient. For example, current research measures instructors' theoretical knowledge through surveys and tests (Boylu, 2019; Çavuşoğlu & Işık, 2021; Karagöl, 2020; Mercan & Göktaş, 2023; Özdemir, 2023; Sertdemir, 2021), but independent studies are needed to address specific stages of assessment tool development. Furthermore, short-term targeted training can offer long-term benefits by addressing specific problem areas.

2. Study Purpose

In language exams, it is important for item writers to know the principles of assessment, write questions considering content validity, and comprehend the assessment objectives of items. An item writer should also know whether an item is appropriate for the target group. In this case, item writers should to be able to determine the difficulty levels of items they produce as well as content validity. It is recommended in the literature to assemble a test with items of varying difficulty which address all ability levels within a target population. In other words, in a test designed to assess reading skills in proficiency and placement exams, the item difficulties of all items should not cluster around a certain difficulty level. For example, a test consisting of only easy or difficult items could jeopardize the validity, reliability, and discriminatory power of the test. Therefore, it is important the item difficulty perceptions of item writers align with the results of item analyses.

Therefore, the aim of this study included, first, to compare the item difficulty perceptions of instructors and item writers in the field of teaching Turkish as a foreign language with the item difficulty indices obtained from item analyses, and second, to determine whether their predictions regarding the item difficulties were correct or not.

Additionally, another aim of this study was the analysis of different factors which influence the consistency of teachers' estimates regarding item difficulty.

If it was determined the item difficulty predictions of instructors were not consistent with the item analyses, therefore, an additional aim of the study, which was dependent on the data results, was to make suggestions on how item writers can improve in this regard.

3. Method

In the study, descriptive survey method based on quantitative variables was used. The descriptive survey method aims to describe an existing situation regarding a specific group through use of a questionnaire. In this survey method, which is frequently used in educational research, data are collected from a specific group at a specific time. Descriptive survey method provides information about the behaviors, ideas, beliefs, knowledge, and so forth of the participants in the survey. Through the collected data – descriptions, comparisons, and classifications – can be made about the questions related to the research (Cohen et al., 2000).

Based on the descriptive survey method, our study aim was to determine the surveyed instructors perceptions of item difficulty. In this context, first, item analyses of tests applied to various groups at different language levels were conducted. Then, a survey was administered to a group of 51 instructors with varying years of experience, levels of education, and education program, asking them to estimate the difficulty level of test items. According to our survey results, instructors' difficulty estimates of the items and difficulty levels of the items according to our analyses were compared, and it was determined to what extent instructors estimates regarding item difficulty levels were compatible.

Considering the literature highlights that the predictions of item writers are more compatible, especially in reading skills and multiple-choice items, we first wanted to determine overall consistency. For this reason, agreement analysis was conducted to provide a more general view to analyze the consistency between item difficulties and instructors' predictions. First, we looked at prediction agreement according to the level for the Common European Framework of Reference for Languages (CEFR). As a result of the low agreement rates in the results obtained from the agreement analysis, we decided to expand the analysis. In addition to agreement analyses, to analyze inter-rater reliability among instructors' predictions, a Fleiss' Kappa analysis was performed (Nichols et al., 2010). As the agreement in this analysis was also low, to understand common patterns and types of errors made by all instructors, we decided to apply confusion matrix for each instructor. Even though confusion matrix is a performance measurement tool for machine learning classification models, the fact that our data was quite suitable for this model persuaded us to consider this analysis method could be useful for our evaluation.

Furthermore, to see whether the instructors' characteristics, which was another aim of our study, had an effect on their predictions, participants were grouped according to various characteristics and ANOVA analyses were performed using SPSS statistical analysis software.

3.1. Sample group

3.1.1. Demographic information

Instructors from various disciplines teach Turkish as a foreign language. The instructors working in TLTC's have also achieved different levels of education such as undergraduate, graduate, and doctorate degrees. In addition to these differences, there are a variety of certificate programs offered for teaching Turkish as a foreign language and some instructors participate in such programs. Thus, to observe whether such differences have an effect on prediction of item difficulty value, in the first stage of the

questionnaire, the instructors were queried about their individual characteristics regarding the field of teaching Turkish as a foreign language.

Figure 1 and Figure 2 illustrate ratios related to participating instructors' educational status along with the university departments from which they graduated.

Figure 1

Instructors' Education Level
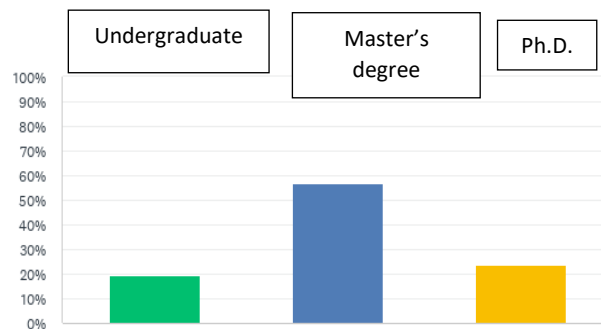
It can be seen in Figure 1 a majority of instructors (approximately 57%) have earned a Master's degree.

Figure 2

Education Programs from which Instructors Graduated
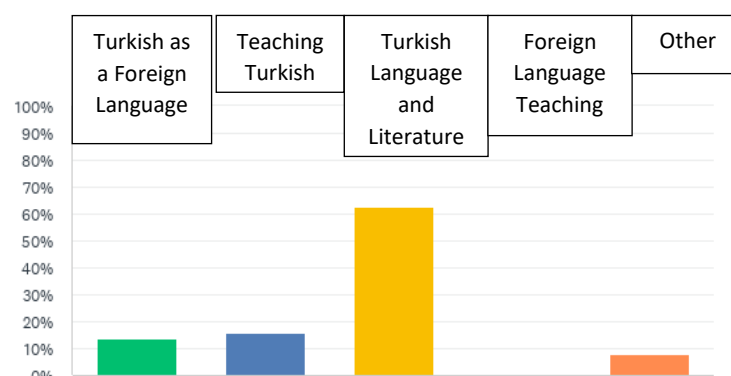
It can be seen in Figure 2 that 14% of participating instructors graduated from a teaching Turkish as a foreign language program, 16% from teaching Turkish, 64% from Turkish language and literature, and 8% from other departments such as Linguistics or elementary school teaching).

Instructors' experience in teaching Turkish as a foreign language was grouped according to 0-1, 2-4, 5-9, and 10 or more years.

Figure 3

Instructors Experience in Years
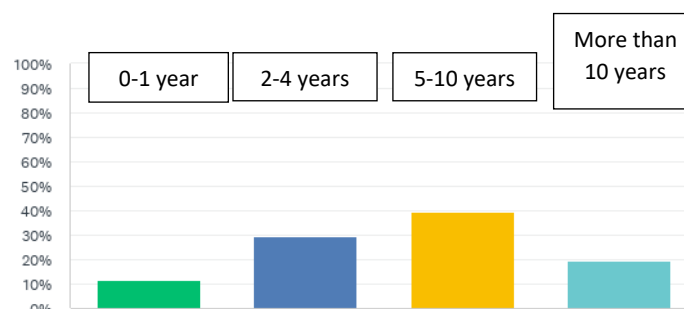
In Figure 3 it can be seen a majority of instructors (approximately 39%) have between two to five years of teaching experience. While, approximately 12% have zero to one year of teaching experience, 29% with two to four years teaching experience, and 20% with 10 or more years teaching experience.

In addition to demographic information presented in Figures 1, 2, and 3, 75% of instructors stated having received training specific to teaching Turkish as a foreign language. Along with this, 85% of instructors stated participating in certificate programs related to teaching Turkish as a foreign language, while 78% stated regularly preparing various exams.

4. Data Collection Methods

For the purpose of this study, two different data collection methodologies were used. The first was to analyze tests utilized in the TLTC. In the beginning stage of this study, the tests were analyzed, while in the second study stage, an item difficulty perception survey was developed and administered to instructors via Survey Monkey. Data obtained from the survey were used to access instructors' demographic information as well as to compare instructors' item difficulty perceptions with difficulty values from the item analysis. Next, the final study stage involved comparing correlations between instructors' item difficulty estimates and the item analyses from different analyses.

4.1. Item analysis

Within the scope of this study, first, item analyses of test sections assessing reading skills. prepared for different levels, were analyzed. For example, at the B1 level, two different tests were used: fill-in-the-blank, to measure grammatical accuracy, and a reading test to assess reading comprehension, which consisted of five multiple-choice (MC) items with each including three options. Similarly, at the B2 level, two tests were used: fill-in-the-blank test, prepared to measure grammatical accuracy, including the appropriate options, and a reading test aimed at assessing reading comprehension skills, which consisted of five multiple-choice (MC) items each with three options. Next, at the C1 level, only a fill-in-the-blank test was utilized to measure grammatical accuracy, vocabulary, and cohesion. Importantly, tests used at a TLTC as part of course achievement exams, was included as part of reading skills assessment. Item analysis for each test was carried out with data from exams administered in the TLTC. Information regarding the number of candidates and item types is presented in Table 1.

Table 1

Information Regarding Tests

Test	Candidate Numbers	Number of Items	Item Types
B1.1	81	10	MC / 3 options
B1.2	81	5	MC / 3 options
B2.1	86	10	MC / 3 options
B2.2	56	5	MC / 3 options
C1.1	59	10	MC / 3 options

Our test data were analyzed through the TAP program with item difficulty, item discrimination and biserial, point biserial correlations calculated for every test. Test reliability was based on the K20 values calculated for each test through TAP. Test reliability, mean item difficulty, and test discrimination rates are presented in Table 2.

Table 2

Information Regarding Reliability, Discrimination, and Mean Difficulty of Tests

Test	Test Reliability [K20(Alpha)]	Test Discrimination (Mean Point Biserial)	Mean Item Difficulty
B1.1	0.672	0.504	0.635
B1.2	0.555	0.600	0.548
B2.1	0.548	0.447	0.653
B2.2	0.585	0.613	0.550
C1	0.638	0.462	0.329

K20 Alpha values in Table 2 show test reliability. Tests with a K20 Alpha value higher than 0.6 have an acceptable reliability level. However, considering the small sample size, all tests in this study had only marginally acceptable reliability with values above or around .60.

Within the scope of this study, the theory proposed by Crocker and Algina (1986) was referred to as a basis for analyzing test items with items with a discrimination index greater than 0.40 determined as good, and items between 0.30 and 0.39 considered acceptable. However, items below 0.30 did not meet discrimination levels. Item difficulty was evaluated between 0 and 1 in the literature and items with 0.50 considered to have moderate difficulty. Importantly, it is generally recommended to include items of moderate difficulty in a test, yet to ensure discrimination, difficult and easy items should also be included in a test. Within the scope of the current study, items between 0.01 and 0.40 were considered difficult, items between 0.41 and 0.60 considered moderate difficulty, and items between 0.61 and 0.99 considered easy. Regardless, discrimination indices should also be considered when evaluating difficulty levels. In particular, items with difficulty values on the borderline such as 0.41 and 0.61 should be examined together with their discrimination indices and data interpreted, along with difficulty classification being made.

4.2. Data collection for item difficulty prediction

In the study's second stage, a questionnaire was applied to instructors working in various institutions for teaching Turkish as a foreign language. In the questionnaire's first stage, instructors were asked various questions such as the department from which they graduated, the degree they earned, amount of teaching experience, and whether they had prepared exams in the centers in which they worked. In the questionnaire's second stage, instructors were asked about their item difficulty evaluations. Texts and items from the reading skills test were sent to the instructors and they were asked to determine item difficulty for each item as "difficult", "moderate", or "easy". The questionnaire prepared for instructors to complete was shared for them to access on "SurveyMonkey". Within the scope of this study, 51 instructors voluntarily completed the survey.

4.3. Analyzing consistency between difficulty level and difficulty prediction

The study's final stage was to evaluate the relationship between instructors' item difficulty perceptions and item analyses. Since there were no studies identified in the literature regarding item difficulty predictions of instructors teaching Turkish as a foreign language, along with no data about the general consistency of instructors in this regard, it was determined to provide a preliminary overview. Therefore, to determine instructors' agreement regarding item difficulties as well as the general view of whether a standard existed or not, we first examined instructors item difficulty predictions on the basis of inter-rater validity.

Consistency between item difficulty analyses and instructor evaluations was measured through method of agreement analysis. For example, the formula from Miles and Huberman (1994) was used in agreement analysis.

((Number of Agree / (Number of Agree + Number of Disagree))*100

The Miles and Huberman formula, generally used for inter-rater reliability, was utilized in this study to calculate agreement among instructor ratings. Instructors' item difficulty ratings were classified by taking into account the rating with the highest percentage. For example, an item rated by instructors as 36% (Low), 56% (Moderate), or 8% (High) was classified as "moderate". Furthermore, item difficulty levels were also coded by classifying them as difficult, moderate, or easy based on the item analysis results.

As a result of the low agreement and lack of consistency in the item difficulty predictions of instructors, a more detailed analysis was necessary. Thus, in addition to agreement analyses, as a means of analyzing

inter-rater reliability among instructors' predictions, a Fleiss' Kappa analysis was performed (Nichols et al., 2010), and a Kappa score of 0.049 found. This score indicated only a slight agreement among instructors' predictions, indicating minimal consensus regarding item difficulty categorization. Additionally, instructors' predictions of item difficulty levels, including low (easy), moderate, and high (difficult) were converted into ordinal values; 1, 2, and, 3, respectively. Difficulty levels of items found through item analyses were also converted into ordinal values, where > .70 was coded as 1 (easy), .70-.30 as 2 (moderate), and < .30 as 3 (difficult).

As a result of low prediction success, it was decided to apply confusion matrix for each instructor. Even though confusion matrix is a performance measurement for machine learning classification models, the fact our data was quite suitable for this model convinced us to incorporate this analysis in the evaluation process. Thus, in this study, focused on instructor perceptions in language assessment, a different analysis methodology was conducted and in effect presented to the field. Indeed, the use of different analysis methods in the field of language assessment can likely increase opportunities for observation as well as to identify unforeseen problems and deficiencies. Based on the results of our use of confusion matrix, we first determined each instructor's item perception success along with interpreting the analysis results. Within the scope of this study, each instructor's matrix results were not shared, but one instructor's results are presented as an example as well as success percentage data of each instructor also presented in Appendix 1.

Next, to obtain detailed analysis data, ANOVA analyses were performed through SPSS. In ANOVA analyses, each instructor's item difficulty prediction (low, moderate, high) as well as item difficulties for the item analyses (easy, moderate, difficult) were coded. A paired sample t-test analysis was also conducted to determine whether prediction success varied at different CEFR levels. Demographic data obtained from the survey regarding instructors were also included in the analyses to understand whether different factors played a role in instructors' success in predicting item difficulty. Additionally, our analysis examined whether the success rates of trainers were related to their experience, level of education, department from which they graduated, having a foreign language teaching certificate, and/or preparing exams.

4.4 Ethical Principles

The ethics committee report for this study was obtained from the Istanbul University Rectorate Social and Human Sciences Research Ethics Committee with the decision dated 11.01.2024 and numbered 2024/19.

5. Findings and Discussion

Different analyses were carried out with differing data at different stages of this study. In the first stage, item analyses of reading tests for different language levels obtained from a TLTC were conducted. There were a different number of candidates and items at each level. In the second stage, the questionnaire percentages for instructors' item difficulty perceptions were determined as well as comparisons made regarding item difficulty values. In the third stage, agreement analysis of the instructors' difficulty ratings and item difficulty ratings were conducted.

5.1. Item analyses of B1, B2, and C1 tests

Following the item analyses, which was the first stage of this study, and conducted through TAP (Test Analysis Program), tables were prepared for analysis of data from each test. Then, these data were interpreted through the classification of theory and difficulty levels determined within the study's scope.

5.1.1. Item analyses of B1 level tests

Table 3

Item Analyses of B1.1 Test

B1.1 / 3 options MC /choosing the correct fill-in-the-blank answer						
Item	Correct Answers	Item difficulty	Discrimination index	High group	Low group	Biserial correlation	Point biserial correlation
Item 1	63	0.78	0.46	25 (0.93)	12 (0.46)	0.57	0.43
Item 2	55	0.68	0.58	25 (0.93)	9 (0.35)	0.56	0.40
Item 3	44	0.54	0.47	22 (0.81)	9 (0.35)	0.47	0.28
Item4	42	0.52	0.73	26 (0.96)	6 (0.23)	0.61	0.45
Item5	50	0.62	0.81	26 (0.96)	4 (0.15)	0.67	0.53
Item6	37	0.46	0.47	18 (0.67)	5 (0.19)	0.43	0.24
Item7	55	0.68	0.50	24 (0.89)	10 (0.38)	0.47	0.29
Item8	60	0.74	0.58	27 (1.00)	11 (0.42)	0.55	0.40
Item9*	54	0.67	0.28	21 (0.78)	13 (0.50)	0.38	0.19
Item10	54	0.67	0.39	25 (0.93)	14 (0.54)	0.35	0.15

In Table 3, it can be seen the discrimination index of Item 9 was below 0.30 and therefore its discrimination was not in the acceptable range. All items except Item 9 had good or acceptable discrimination indices. Thus, it can be seen from the analysis that Items 4 and 6 had moderate difficulty and all other items had low difficulty levels, that is, they should be considered as easy.

Table 4

Item Analyses of B1.2 Test

B1.2 / 3 options MC / reading comprehension						
Item	Correct Answers	Item difficulty	Discrimination index	High group	Low group	Biserial correlation	Point biserial correlation
Item 1	34	0.42	0.90	23 (0.96)	2 (0.06)	0.69	0.44
Item 2	53	0.65	0.54	22 (0.92)	12 (0.38)	0.57	0.28
Item 3	58	0.72	0.49	23 (0.96)	15(0.47)	0.59	0.33
Item 4	30	0.37	0.72	18 (0.75)	1 (0.03)	0.64	0.36
Item 5	47	0.58	0.56	21 (0.88)	10 (0.31)	0.51	0.19

In Table 4, it is seen the discrimination indices for all items in the test were valid. In terms of difficulty level, it was determined Item 4 was difficult, Item 1 was moderate (close to difficult), Item 5 was moderate, and Items 2 and 3 were easy.

5.1.2. Item analyses of B2 level tests

Table 5

Item Analyses of B2.1 Test

B2.1 / 3 options MC /choosing the correct fill-in-the-blank answer						
Item	Correct Answers	Item difficulty	Discrimination index	High group	Low group	Biserial correlation	Point biserial correlation
Item 1	65	0.76	0.32	30 (0.88)	14 (0.56)	0.36	0.15
Item 2	66	0.77	0.41	33 (0.97)	14 (0.56)	0.38	0.18
Item 3	65	0.76	0.42	32 (0.94)	13 (0.52)	0.47	0.27
Item 4	39	0.45	0.59	27 (0.79)	5 (0.20)	0.52	0.31
Item 5	72	0.84	0.34	32 (0.94)	15 (0.60)	0.48	0.32
Item 6	58	0.67	0.61	29 (0.85)	6 (0.24)	0.58	0.38
Item 7	38	0.44	0.43	20 (0.59)	4 (0.16)	0.39	0.15
Item 8	41	0.48	0.47	24 (0.71)	6 (0.24)	0.40	0.17
Item 9	68	0.79	0.46	32 (0.94)	12 (0.48)	0.47	0.29
Item 10	51	0.59	0.47	27 (0.79)	8 (0.32)	0.38	0.14

The discrimination indices for all items in Test 1 at the B2 level were above the acceptability value. In terms of difficulty level, Items 1, 2, 3, 5, 6, 9, and 10 were easy, while Items 4, 7, and 8 were of moderate

difficulty. Although there were no items considered difficult within the test, Item 6 had a discrimination index of 0.61, and Item 4 had a discrimination index of 0.59 which stood out in terms of low group and high group correct answer rates.

Table 6

Item Analyses of B2.2 Test

B2.2 / 3 options MC / reading comprehension							
Item	Correct Answers	Item difficulty	Discrimination index	High group	Low group	Biserial correlation	Point biserial correlation
Item 1	37	0.66	0.59	20 (1.00)	11 (0.41)	0.63	0.38
Item 2	23	0.41	0.74	17 (0.85)	3 (0.11)	0.69	0.45
Item 3	31	0.55	0.73	19 (0.95)	6 (0.22)	0.66	0.41
Item 4	29	0.52	0.55	17 (0.85)	8 (0.30)	0.55	0.26
Item 5	34	0.61	0.43	16 (0.80)	10 (0.37)	0.52	0.23

All test items in Table 6 had acceptable discrimination indices. Although the difficulty level of Item 2 was considered in the intermediate group, it could be accepted as difficult due to its high discrimination index. Item 3 could be evaluated similarly and accepted as difficult. On the other hand, Item 4, which had a difficulty level of 0.52, was considered to be of moderate difficulty due to its discrimination index being 0.55.

5.1.3. Item analyses of C1 level tests

Table 7

Item Analyses of C1 Test

C1 / 3 options MC / choosing the correct fill-in-the-blank answer							
Item	Correct Answers	Item difficulty	Discrimination index	High group	Low group	Biserial correlation	Point biserial correlation
Item 1	56	0.95	0.12	28 (1.00)	15 (0.88)	0.42	0.29
Item 2	51	0.86	0.47	28 (1.00)	9 (0.53)	0.62	0.46
Item 3	53	0.90	0.24	28 (1.00)	13 (0.76)	0.36	0.17
Item 4	47	0.80	0.59	28 (1.00)	7 (0.41)	0.57	0.35
Item 5	50	0.85	0.41	28 (1.00)	10 (0.59)	0.59	0.40
Item 6*	58	0.98	0.00	28 (1.00)	17 (1.00)	-0.02	-0.10
Item 7	54	0.92	0.29	28 (1.00)	12 (0.71)	0.48	0.32
Item 8	51	0.86	0.47	28 (1.00)	9 (0.53)	0.69	0.54
Item 9	49	0.83	0.35	28 (1.00)	11 (0.65)	0.38	0.15
Item 10	51	0.86	0.35	28 (1.00)	11 (0.65)	0.53	0.34

In Table 7, it can be seen the difficulty level of all items at the C1 level were easy. Also, it was recognized the discrimination indices of Items 1, 3, and 7 were outside the acceptable level as well as Item 6 had no discrimination. In fact, it can be seen the item worked in reverse when the biserial correlation and point biserial correlation of Item 6 were examined. However, considering the fact 58 out of 59 candidates who participated in the test answered the item correctly, the assessment that the item worked in reverse should not be accepted as valid. As a matter of fact, only one candidate answered the item incorrectly. Considering the difficulty level of all items in the test was low and the discrimination indices were generally below the acceptable limit, it was not possible to distinguish between the upper and lower groups. In short, it can be interpreted the items in this test did not reach the measurement target. As a matter of fact, the use of a measurement tool in achievement tests, in which all items are easy, does not serve any measurement purpose. However, since the purpose of this study was to determine instructors' perception of item difficulty, there was a need to evaluate test items with low discrimination. For this reason, the C1 test was also included in the study and the instructors' evaluations of test items also examined.

Item analyses for B1, B2, and C1 level tests revealed variations in difficulty and discrimination indices. While the B1 and B2 tests performed within acceptable parameters, with most items being easy but

displaying appropriate discrimination, the C1 level test showed significant inadequacies. Additionally, items in the C1 test were largely too easy, and several failed to meet discrimination standards, reflecting broader issues in language testing practices as identified by Özdemir and Eke (2023), Sertdemir (2021), and Şimşek (2016). These findings highlighted a need for improvement in item-writing practices, particularly for higher-level assessments, to ensure accurate measurement of student abilities, especially in areas requiring higher-order cognitive skills (Aydemir & Çiftçi, 2008; Oktay, 2015; Özcan & Akçan, 2010).

5.2. Findings related to instructors' perceptions of item difficulty

After the item analyses of the achievement tests analysed within the scope of this study was evaluated in the light of certain theories, an item difficulty rating questionnaire was prepared for instructors in the second study stage. In the survey, text and item levels were specified and instructors were asked to determine item difficulty as "low", "moderate", or "high" for each item. Data were collected through the "SurveyMonkey" website.

Instructors' item difficulty perceptions were analysed through the "SurveyMonkey" survey application. In the analysis, instructors' evaluations for each item were ranked according to all participating instructors. To determine agreement of instructors' perceptions with item difficulty levels, the percentages showing instructors' perceptions of difficulty levels for items in each test along with item difficulty values obtained from the item difficulty analysis were compared.

5.2.1. Comparison of instructors' perceptions of difficulty and item difficulty values for the reading test B1.1

Item analysis of the first reading test at the B1 level was conducted regarding the results of 81 candidates. The test consisted of 10 MC items with three options. Candidates were asked to choose the appropriate answers for the fill-in-the-blanks in a 450-word text. Difficulty values for the items and instructors' ratings are presented in Table 8.

Table 8

Instructors' Item Difficulty Perception Ratings and Item Difficulty Analyses for B1.1

	Rates of instructors' assessment of item difficult perception	Item difficulty
Item 1	34% Low 58% Moderate 12% High	0.78
Item2	32% Low 56% Moderate 12% High	0.68
Item3	22% Low 48% Moderate 30% High	0.54
Item4	30% Low 46% Moderate 24% High	0.52
Item5	12% Low 62% Moderate 26% High	0.62
Item6	26% Low 50% Moderate 24% High	0.46
Item7	46% Low 36% Moderate 18% High	0.68
Item8	26% Low 64% Moderate 10% High	0.74

Table 8(Continued)

Item9	44% Low	0.67
	54% Moderate	
	2% High	
Item10	26% Low	0.67
	50% Moderate	
	24% High	

5.2.2. Comparison of instructors' perceptions of difficulty and item difficulty values for reading test B1.2

Item analysis for the second reading test at the B1 level was conducted regarding the results of 81 candidates. In this test, there was a text consisting of 489 words and five MC items with three options for reading comprehension.

Table 9

Instructors' Item Difficulty Perception Rates and Item Difficulty Analyses for B1.2

	Rates of instructors' assessment of item difficulty perception	Item difficulty
Item 1	48% Low	0.42
	36% M%oderate	
	16% High	
Item 2	44% Low	0.65
	44% Moderate	
	12% High	
Item 3	62% Low	0.72
	26% Moderate	
	12% High	
Item 4	14% Low	0.37
	52% Moderate	
	34% High	
Item 5	54% Low	0.58
	40% Moderate	
	6% High	

Item 4 in the B1.2 test differed from other items with a difficulty level of 0.37. A majority of instructors rated Item 4 as the most difficult test item, at Moderate difficulty. Also, a majority of instructors considered Item 5, which had Moderate difficulty for analysis data, as an easy item.

5.2.3. Comparison of instructors' perceptions of difficulty and item difficulty values for reading test B2.1

The first reading test at the B2 level consisted of 10 fill-in-the-blanks in a text consisting of 56 words and 10 MC items with three options where the appropriate option was marked. Item analysis of the test was conducted with the data regarding 86 candidates.

Table 10

Instructors' Item Difficulty Perception Rates and Item Difficulty Analyses for B2.1

	Rates of instructors' assessment of item difficulty perception	Item difficulty
Item 1	36% Low 54% Moderate 10% High	0.76
Item 2	32% Low 56% Moderate 12% High	0.77
Item 3	40% Low 48% Moderate 12% High	0.76
Item 4	16% Low 48% Moderate 36% High	0.45
Item 5	36% Low 52% Moderate 12% High	0.84
Item 6	34% Low 56% Moderate 10% High	0.67
Item 7	10% Low 64% Moderate 26% High	0.44
Item 8	18% Low 46% Moderate 36% High	0.48
Item 9	16% Low 54% Moderate 30% High	0.79
Item 10	32% Low 50% Moderate 18% High	0.59

5.2.4. Comparison of instructors' perceptions of difficulty and item difficulty values for reading test B2.2

In the second reading test at the B2 level, data from 56 candidates were used for item analysis. In the test, there were five MC items with three options for reading comprehension regarding a text of 516 words.

Table 11

Instructors' Item Difficulty Perception Rates and Item Difficulty Analyses for B2.1

	Rates of instructors' assessment of item difficulty perception	Item difficulty
Item 1	48% Low 34% Moderate 18% High	0.66
Item 2	48% Low 38% Moderate 14% High	0.41
Item 3	34% Low 52% Moderate 14% High	0.55
Item 4	36% Low 46% Moderate 18% High	0.52
Item 5	48% Low 36% Moderate 16% High	0.61

Although Item 2 in the B2.1 test was the item with the highest difficulty level, it was largely evaluated as an Easy or Moderate difficulty item by instructors. A similar situation was also observed for Item 5.

5.2.5. Comparison of instructors' perceptions of difficulty and item difficulty values for reading test C1

Item analysis of one C1 level test provided usable data for our study. Item analysis of the second test shared by the TLTC could not be conducted due to all participants answering the items correctly. Therefore, item analysis could only be performed on the first test. Thus, data from 59 candidates were used for item analysis. The test included a reading text of 494 words and 10 MC items with three options with the instruction to fill in 10 blanks within the text for the right option.

Table 12

Instructors' Item Difficulty Perception Rates and Item Difficulty Analyses for C1

	Rates of instructors' assessment of item difficulty perception	Item difficulty
Item 1	62% Low 34% Moderate 4% High	0.95
Item2	64% Low 30% Moderate 6% High	0.86
Item3	26% Low 56% Moderate 18% High	0.90
Item4	10% Low 68% Moderate 22% High	0.80
Item5	12% Low 66% Moderate 22% High	0.85
Item6	68% Low 30% Moderate 2% High	0.98
Item7	32% Low 54% Moderate 14% High	0.92
Item8	44% Low 38% Moderate 18% High	0.86
Item9	20% Low 48% Moderate 32% High	0.83
Item10	36% Low 56% Moderate 8% High	0.86

The most striking data in Table 12 highlights that Item 3, which had a very low difficulty value, was accepted by 56% of instructors at a moderate level. It was previously reported the discrimination index of Item 3 was also outside the acceptable value for item analysis. A similar difference was observed in Items 4, 5, 7, 9, and 10. In Item 8, although the percentages were close, 38% of instructors thought the item was of moderate difficulty for an item with very low difficulty level, and as a result, should be considered easy. The rate of instructors who found the difficulty level low was 44%. In short, instructors

almost equally chose easy and moderate difficulty levels for this item. Additionally, data in the C1 test clearly showed instructors' item difficulty perceptions were not compatible with the item analyses. In other tests, this discrepancy was lower than in the C1 test, however, it can still be seen instructors' perceptions of item difficulty values were mostly inconsistent with the analyses.

Previous research also indicates significant gaps in the assessment and evaluation competencies of instructors. According to Ustabulut (2021) as well as Erdoğdu and Kurt (2012), instructors' ability to analyse exam results and make evaluations aligned with learning objectives is at a moderate level. On the other hand, Yıldız and Tepeli (2014) found instructors demonstrate high competency in applying contemporary assessment and evaluation methods. However, studies by Hatipoğlu (2015), Mede and Atay (2017), and Ölmezer-Öztürk and Aydın (2018) revealed instructors lack sufficient knowledge in the field of assessment and evaluation. Similar findings are observed in international studies with Ahmadi and Ketabi (2020), Bahtiar and Purnawarman (2020), Fitriyah et al. (2022), and Latif (2021) reporting instructors feel inadequate in the area of foreign language assessment literacy. Additionally, Bøhn and Tsagari (2021), Firoozi et al. (2019), Liu and Li (2020), Razavipour and Rezagah (2018), and Sultana (2019) emphasized instructors lack necessary skills in exam preparation and learner evaluation.

5.3. Analyses for instructors' predictions

In the final study stage, agreement analysis was conducted between the instructors' item difficulty perceptions and item difficulty analyses. By reaching the quantitative results for evaluations made in both stages with agreement analysis, the rate of difference between instructors' item difficulty perceptions and item analyses was determined.

Percentages of agreement determined after applying the formula are presented in Table 13.

Table 13

Instructors' Item Difficulty Perception and Item Difficulty Analysis Compatibility

Test	Agreement rate	Agreement value
B1.1	40%	Low
B1.2	40%	Low
B2.1	40%	Low
B2.2	80%	High
C1	40%	Low

As can be seen in Table 13, instructors' item difficulty perceptions in general along with difficulty values in the item analyses were not consistent. Thus, the percentage agreement for each instructor was calculated to determine their success rate in terms of predicting the difficulty level of an item. As a result, the percentage of correctly predicted items for each instructor was calculated.

The percentage of correct predictions by each instructor ranged from 17.5% to 72.5%. Variation in success rates indicated a significant disparity in instructors' ability to accurately predict item difficulty. For example, some instructors were able to match actual difficulty levels of items more accurately than others. Instructors with success rates above 50% showed relatively high predictive accuracy. Therefore, these instructors might possess better intuition or experience in accurately judging item difficulty. For instance, instructors with success rates around 72.5% might be leveraging their extensive experience and/or specific training in assessment. Whereas instructors with success rates below 30% struggled to accurately predict item difficulty. This could be due to various factors, such as less experience, lack of specific training in item difficulty assessment, and/or differing perceptions of what constitutes item difficulty.

The next step involved development of a confusion matrix to understand the type of errors made by each instructor. A confusion matrix allows for comprehensive evaluation of how well a model performs as well as where it might go wrong (Witten et al., 2005). In our study, in the context of evaluating

instructors' predictions of item difficulty, using a confusion matrix aided in identifying specific patterns regarding their predictions. The matrix included:

- True Positives (TP): Correctly predicted difficulty levels.
- False Positives (FP): Predicted a higher difficulty level than actual.
- False Negatives (FN): Predicted a lower difficulty level than actual.

For instance, a moderate item predicted as moderate was a true positive. A false positive could be exemplified by a moderate item predicted as easy. Whereas a moderate item predicted as difficult was an example of a false negative. Considering Instructor 1 as an example, for easy items, this instructor correctly identified 18 as easy, but incorrectly classified eight as medium and one as difficult. For medium items, Instructor 1 correctly predicted five as medium, but mistakenly labelled six as easy and one as difficult. When it came to difficult items, Instructor 1 correctly identified none, misclassifying one as easy and none as medium.

Table 14

Confusion matrix sample (for Instructor 1)

Predicted	Actual	Count	Instructor ID
Easy	Easy	18	1
Easy	Moderate	6	1
Easy	Difficult	1	1
Moderate	Easy	8	1
Moderate	Moderate	5	1
Moderate	Difficult	0	1
Difficult	Easy	1	1
Difficult	Moderate	1	1
Difficult	Difficult	0	1

Next, to understand common patterns and types of errors made by all instructors, we aggregated the confusion matrices for each instructor. The aggregated confusion matrix revealed several key insights. For easy items, 37.18% were correctly predicted as easy, 48.44% were incorrectly predicted as medium, and 14.38% incorrectly predicted as difficult. For medium items, 46.90% were correctly predicted as medium, while 31.54% were incorrectly predicted as easy, and 21.57% incorrectly predicted as difficult. For difficult items, only 33.33% were correctly predicted as difficult, with 50.98% being incorrectly predicted as medium, and 15.69% as easy.[1]

The most frequent misclassification occurred between easy and medium items. For example, instructors tended to overestimate the difficulty of easy items, predicting them as medium, and underestimating the difficulty of medium items, predicting them as easy. This indicated a general tendency to perceive items within a narrower range of difficulty, often defaulting to a medium rating.

Moreover, difficult items were often underestimated, with a significant portion being predicted as medium and some even as easy. These suggested instructors had a challenging time accurately identifying items as difficult, likely due to a lack of clear distinguishing characteristics for such items.

---

[1] Success percentages for each instructor are presented in Appendix 1.

Medium items were somewhat more accurately predicted compared to easy and difficult items, with 46.90% correct predictions. However, there was still notable confusion, especially with predictions leaning towards easy.

The fact the prediction agreement percentages of instructors were low in general, brought to mind the question of whether experience, program graduation, education level, and so forth had some positive effect on item difficulty perception.

Instructors' predictions of item difficulty were compared to actual difficulty levels to calculate success rates. Prediction success rates were calculated as the percentage agreement between instructors' predictions and actual difficulty levels. To understand which demographic variables affected these success rates, group differences were analysed through independent t-tests and ANOVA (based on the number of variable categories) for each CEFR level separately (B1, B2, and C1). Descriptive statistics for instructors' success rates at different CEFR levels are presented in Table 15.

Table 15

Instructors' Success Rates at Different CEFR Levels

CEFR Level	N	Mean	Std. Deviation
B1	40	0.512	0.129
B2	40	0.458	0.136
C1	40	0.372	0.147

Assessment training was determined to be the only binary variable with significant group difference in instructors' prediction success rate at the B2 level. The t-test results also showed a significant difference $(t(20.94) = -2.682, p = 0.0131)$, indicating instructors with assessment training had significantly higher success rates.

To investigate the impact of various demographic variables on instructors' success rates in predicting item difficulty at different CEFR levels, a series of ANOVA tests were conducted. The variables examined included education level, education department, and teaching experience. Multifactorial ANOVA was also performed to explore possible interactions which effected item difficulty prediction success rates for instructors, but no significant interaction was found.

Furthermore, ANOVA results for the B1 level indicated none of the demographic variables examined (i.e., education, department, experience) had a significant effect on instructors' success rates. The lack of significant findings suggested these factors did not influence instructors' ability to predict item difficulty at the B1 level.

Similarly, for the B2 level, ANOVA results showed no significant effect of demographic variables (i.e., education, department, experience) on instructors' success rates. This indicated these factors did not significantly impact instructors' accuracy in predicting item difficulty at the B2 level.

Contrary to expectations, ANOVA analysis for the B1 and B2 levels did not reveal any significant effects of the demographic variables examined. This suggested factors such as education level, department, and teaching experience did not substantially influence instructors' ability to predict item difficulty for lower and intermediate proficiency levels.

For the C1 level, ANOVA results revealed a significant effect of 'Experience' on success rates $(F(3, 38) = 3.7233, p = 0.0193)$. This suggested teaching experience played a crucial role in predicting the difficulty of C1 level items. To further explore specific group differences within the 'Experience' variable, post-hoc comparisons using the Tukey HSD test were performed. The Tukey HSD test results indicated instructors with more than 10 years of experience had significantly higher success rates compared to

those with 2-4 years of experience. Considering almost all of the items at the C1 level were classified as easy according to the item analysis results, it can be concluded this achievement difference was not important. Because at this point, guessing factors can be considered high. Moreover, the fact there was no significant effect of experience on prediction success at other levels supported this finding.

Therefore, our study's findings indicated that demographic factors such as education level, education department, and teaching experience did not significantly affect instructors' ability to accurately predict item difficulty at the B1 and B2 CEFR levels, aligning with Tao's (2014) assertion that formal education, including undergraduate and postgraduate degrees, may not substantially enhance teachers' competencies in language assessment. This highlighted the inadequacy of short-term academic courses, as noted by Sultana (2019) as well as Yan and Fan (2021) in fostering a deep understanding of assessment. The current study also reinforced Tao's (2014) argument that teaching experience alone is insufficient for developing assessment literacy, also supporting Fitriyah, Massitoh, and Widiati (2022) which called for ongoing and targeted professional development. While this study determined teaching experience had a significant impact on predicting item difficulty at the C1 level, particularly for those with over 10 years of experience, the influence of guessing due to the ease of items should be considered. This was consistent with Levi and Inbar-Lourie (2020) who concluded theoretical knowledge without practical application may not lead to better performance. Similarly, Ölmezer-Öztürk and Aydın (2018) found even experienced instructors require structured, experiential learning opportunities to improve their assessment skills.

5. Conclusion

This study highlighted critical gaps in instructors' ability to predict item difficulty within the context of teaching Turkish as a foreign language. Despite the availability of research regarding the competencies of educators, there remained a scarcity of descriptive studies focused on identifying specific deficiencies, especially in the realm of assessment and evaluation. This study aimed to fill this gap by investigating the accuracy of instructors' perceptions of item difficulty along with the implications for validity and reliability of exams.

Our findings revealed instructors often struggled with correctly assessing item difficulty, which had serious implications on test validity. This result aligned with previous research by Alderson (1993) and Bachman (2002), who both observed significant challenges among even experienced educators in predicting the difficulty levels of test items. These studies further demonstrated instructors, particularly those with extensive experience in exam preparation, may overestimate their ability to accurately gauge item difficulty. In fact, more experienced instructors tended to have lower agreement percentages in their item difficulty predictions. Similarly, Shohamy (1984), observed response format familiarity, particularly with multiple-choice items, can lead to misjudgements regarding item difficulty. This suggested experience alone is insufficient for accurate item difficulty assessment without necessary training and established analytical skills.

Moreover, correct determination of item difficulty is crucial for ensuring items are appropriately classified according to a learners' language proficiency level. Inaccuracies in this process can compromise both exam validity and reliability. For instance, instructors who misclassify difficult items as "easy" raise questions about their ability to create level-appropriate assessments. This issue is further supported by Bejar (1983), who demonstrated the use of anchor-based methods and structured training can significantly enhance accuracy of item difficulty predictions. As a result, these findings underscored the need for more comprehensive training programs focused on item analysis and test construction.

Furthermore, this study drew attention to the need for improving instructors' understanding of item characteristics, particularly those which influence difficulty levels. Training focused on recognizing features of challenging items can help instructors more accurately predict item difficulty, leading to

improved item classification, and ultimately more valid and reliable assessments. Freedle and Kostin (1999) emphasized topic familiarity plays a significant role in item difficulty predictions, while Rupp, Garcia, and Jamieson (2001) highlighted the importance of lexical knowledge. Incorporating these elements into professional development programs can greatly benefit instructors' assessment literacy.

Therefore, within the scope of our study, the following suggestions were made:

-Teachers should undergo specialized training aimed at enhancing their proficiency in item construction, with particular emphasis on developing items that measure higher-order cognitive skills.

-Teacher education programs at universities should incorporate a greater number of courses regarding assessment and evaluation, supplemented with practical, hands-on learning opportunities.

-Mentorship and support systems should be established, whereby experienced educators provide guidance and feedback to less experienced teachers, particularly in the context of test item development.

-The process of item construction should be standardized through the implementation of structured guidelines and templates, and all test items should be subject to pilot testing prior to formal administration.

-Regular reviews and revisions of test items should be conducted, particularly for items with low discrimination indices, to ensure assessment reliability and validity.

-Mechanisms for structured feedback should be developed to improve teachers' accuracy in predicting item difficulty, along with more advanced training focused on item analysis and difficulty estimation.

-In high-stakes examinations, particularly at the C1 level and above, a stronger focus should be placed on the development of test items which assess advanced cognitive abilities, in alignment with Bloom's Revised Taxonomy.

-The implementation of mandatory pilot testing and comprehensive data analysis for all assessments should be enforced to ensure the validity and discriminatory power of test items prior to their final administration.

References

Ahmadi, M. R. S., & Ketabi, S. (2020). Features of language assessment literacy in Iranian English language teachers' perceptions and practices. *Journal of Teaching Language Skills*, *38*(1), 191-223. https://doi.org/10.22099/jtls.2020.34843.2739

Alderson, C. (1993). Judgments in language testing. In D. Douglas & C. Chapelle (Eds.), *A new decade in language testing* (pp. 46-57). TESOL.

Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge University Press.

Altıntaş, N. (2022). *Yabancı dil olarak Türkçe öğreten öğretim elemanlarının sınav hazırlama ve ölçme değerlendirme yeterlik algıları [Perceptions of Exam Preparation and Assessment Competencies Among Instructors Teaching Turkish as a Foreign Language]* (Thesis Number: 747926) [Master Thesis, Dokuz Eylül University]. Turkish Council of Higher Education Thesis Center.

Aydemir, P. D. Y., & Çiftçi, Y. Ö. (2008). A study on the questioning skills of literature teacher candidates. *Van Yüzüncü Yıl University Journal of Education Faculty*, *5*(2), 103-115. Retrieved from https://dergipark.org.tr/en/pub/yyuefd/issue/13714/166035

Bachman, L. F. (2002). Some reflections on task-based language performance assessment. *Language Testing*, *19*(4), 453-476. https://doi.org/10.1191/0265532202lt240oa

Bahtiar, I., & Purnawarman, P. (2020). Investigating English teachers' comprehension in language assessment literacy (LAL). *Advances in Social Science, Education and Humanities Research*, *508*, 303-310. https://doi.org/10.2991/assehr.k.201214.253

Bejar, I. I. (1983). Subject matter experts' assessment of item statistics. *Applied Psychological Measurement*, *7*(3), 303-310. https://doi.org/10.1177/014662168300700305

Bøhn, H., & Tsagari, D. (2021). Teacher educators' conceptions of language assessment literacy in Norway. *Journal of Language Teaching and Research*, *12*(2), 222-233. https://doi.org/10.17507/jltr.1202.02

Boylu, E. (2019). *Yabancılara Türkçe öğretiminde ölçme değerlendirme uygulamaları ve standart oluşturma. [Measurement and evaluation practices and standardization in teaching Turkish to foreigners]* (Thesis Number: 542435) [Doctoral Dissertation, Çanakkale Onsekiz Mart University]. Turkish Council of Higher Education Thesis Center.

Çavuşoğlu, R., & Işık, A. D. (2021). Assessment and evaluation process of Turkish language teaching centers (TÖMER). *The Journal of Limitless Education and Research*, 6(2), 291-315. https://doi.org/10.29250/sead.958711

Cohen, L., Manion, L., & Morrison, K. (2000). *Research methods in education (5th ed.)*. Routledge.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Cengage Learning.

Erdoğdu, M. Y., & Kurt, F. (2012). Investigation of teachers' perception of their competencies in assessment and evaluation in terms of certain variables. *Electronic Journal of Education Sciences*, *1*(2), 23-36. Retrieved from https://dergipark.org.tr/en/pub/ejedus/issue/15938/167586

Firoozi, T., Razavipour, K., & Ahmadi, A. (2019). The language assessment literacy needs of Iranian EFL teachers with a focus on reformed assessment policies. *Language Testing in Asia*, *9*(1), 1-12. https://doi.org/10.1186/s40468-019-0078-7

Fergadiotis, G., Swiderski, A., & Hula, W. D. (2019). Predicting confrontation naming item difficulty. *Aphasiology*, *33*(6), 689–709. https://doi.org/10.1080/02687038.2018.1495310

Fitriyah, I., Massitoh, F., & Widiati, U. (2022). Classroom-based language assessment literacy and professional development need between novice and experienced EFL teachers. *Indonesian Journal of Applied Linguistics*, *12*(1), 124-134. https://doi.org/10.17509/ijal.v12i1.46539

Freedle, R., & Kostin, I. (1999). Does the text matter in a multiple-choice test of comprehension? The case for construct validity of the TOEFL's minitalks. *Language Testing*, *16*(1), 2-32. https://doi.org/10.1177/026553229901600101

Gedik, E. (2017). *Yabancı dil olarak Türkçe öğretiminde ölçme ve değerlendirme [Assessment and Evaluation in Teaching Turkish as a Foreign Language]* (Thesis Number: 458489) [Master Thesis, Istanbul Arel University]. Turkish Council of Higher Education Thesis Center.

Hambleton, R. K., & Jirka, S. J. (2006). Anchor-based methods for judgmentally estimating item statistics. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 399-420). Lawrence Erlbaum Associates.

Hatipoğlu, Ç. (2015). Language assessment in English teacher education programs in Turkey. *Journal of Language Testing*, *32*(2), 243-265. https://doi.org/10.1177/0265532214565635

Işıkoğlu, M. (2015). *Yabancı dil olarak Türkçe öğretiminde kullanılan yeterlik sınavlarının madde yazımı bakımından incelenmesi (Mersin ve Sakarya üniversiteleri örneği) [Analysis of proficiency exams developed for teaching Turkish as a foreign language in terms of item writing: Samples of Mersin and Sakarya Universities]* (Thesis Number: 394791) [Doctoral Dissertation, Gazi University]. Turkish Council of Higher Education Thesis Center.

Karagöl, E. (2020). Proficiency exams in teaching Turkish as a foreign language in TÖMER (Turkish and foreign languages research and application centers). *Journal of Language and Linguistic Studies*, *16*(2), 930-947. https://doi.org/10.17263/jlls.759347

Kostin, I. (2004). Exploring item characteristics that are related to the difficulty of TOEFL dialogue items. *ETS Research Report Series*, *2004*(1), i-71. https://doi.org/10.1002/j.2333-8504.2004.tb01938.x

Kutlu, Ö., Doğan, C. D., & Karakaya, İ. (2010). *Öğrenci başarısının belirlenmesi: Performansa ve portfolyoya dayalı durum belirleme [Assessing Student Achievement: Performance and Portfolio-Based Evaluation] (3rd ed.)*. Pegem Akademi.

Latif, M. W. (2021). Exploring tertiary the EFL practitioners' knowledge base component of assessment literacy: Implications for teacher professional development. *Language Testing in Asia*, *11*(1), 1-22. https://doi.org/10.1186/s40468-021-00130-9

Levi, T., & Inbar, O. (2019). Assessment literacy or language assessment literacy: Learning from the teachers. *Language Assessment Quarterly*, *17*(3), 1-15. https://doi.org/10.1080/15434303.2019.1692347

Liu, J., & Li, X. (2020). Assessing young English learners: Language assessment literacy of Chinese primary school English teachers. *International Journal of TESOL Studies*, *2*, 36-49. https://doi.org/10.46451/ijts.2020.12.05

Mercan, Ö., & Göktaş, B. (2023). Türkçenin yabancı dil olarak öğretimi sertifika programları ile CELTA'nın karşılaştırılması: Bir program önerisi [A Comparison of Turkish as a Foreign

Language Certification Programs and CELTA: A Program Proposal]. *Cumhuriyet International Journal of Education*, *12*(3), 715-731. https://doi.org/10.30703/cije.1258699

Mede, E., & Atay, D. (2017). English language teachers' assessment literacy: The Turkish context. *Dil Dergisi*, *168*(1), 1-5. Retrieved from https://www.ceeol.com/search/article-detail?id=548906

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded source book (2nd ed.)*. Sage Publications.

Nichols, T. R., Wisner, P. M., Cripe, G., & Gulabchand, L. (2010). Putting the kappa statistic to use. *The Quality Assurance Journal*, *13*(3-4), 57-61. https://doi.org/10.1002/qaj.481

Oktay, M. R. (2015). *An analysis of sub-questions in Turkish textbooks used in teaching Turkish as a foreign language in terms of cognitive levels in Bloom's Taxonomy* (Thesis Number: 463400) [Master Thesis, Başkent University]. Turkish Council of Higher Education Thesis Center.

Ölmezer Öztürk, E., & Aydın, B. (2018). Developing and validating language assessment knowledge scale (LAKS) and exploring the assessment knowledge of EFL teachers. *Hacettepe University Journal of Education*, *34*(3), 602-620. https://doi.org/10.16986/HUJE.2018043465

Özcan, S., & Akçan, K. (2010). Investigation of questions prepared by science teacher candidates in terms of content and Bloom's Taxonomy. *Kastamonu Journal of Education*, *18*(1), 323-330. Retrieved from https://dergipark.org.tr/en/pub/kefdergi/issue/49066/626064

Özdemir, S. (2023). Yabancı dil olarak Türkçe öğretiminde ölçme ve değerlendirmeye yönelik araştırmaların eğilimleri [Trends in Research on Assessment and Evaluation in Teaching Turkish as a Foreign Language]. *İnönü Üniversitesi Eğitim Fakültesi Dergisi*, *24*(1), 537-559. https://doi.org/10.17679/inuefd.1071671

Özdemir, S., & Eke, H. (2023). Yabancılara Türkçe öğretiminde uygulanan A1 ve A2 kur sınavlarının madde yazma ilkeleri açısından incelenmesi [An Examination of A1 and A2 Level Exams in Teaching Turkish to Foreigners in Terms of Item Writing Principles]. *Trakya Eğitim Dergisi*, *13*(1), 365-380. https://doi.org/10.24315/tred.700445

Razavipour, K., & Rezagah, K. (2018). Language assessment in the new English curriculum in Iran: Managerial, institutional, and professional barriers. *Language Testing in Asia*, *8*(1), 1-12. https://doi.org/10.1186/s40468-018-0061-8

Rupp, A. A., Garcia, P., & Jamieson, J. (2001). Combining multiple regression and CART to understand item difficulty in second language reading and listening comprehension tests. *International Journal of Testing*, *1*(3-4), 185-216. https://doi.org/10.1080/15305058.2001.9669479

Sertdemir, E. (2021). *An analysis of Turkish proficiency exams used in teaching Turkish as a foreign language according to the Revised Bloom's Taxonomy* (Thesis Number: 652111) [Master Thesis, Çanakkale Onsekiz Mart University]. Turkish Council of Higher Education Thesis Center.

Shohamy, E. (1984). Does the testing method make a difference? The case of reading comprehension. *Language Testing*, *1*(2), 147-170. https://doi.org/10.1177/026553228400100203

Şimşek, A. (2016). A comparative analysis of common mistakes in achievement tests prepared by school teachers and corporate trainers. *European Journal of Science and Mathematics Education*, *4*(4), 477-489. https://doi.org/10.30964/ejsme.v4i4.467

Tao, N. (2014). *Development and validation of classroom assessment literacy scales: English as a foreign language (EFL) teachers in a Cambodian higher education setting* (Unpublished doctoral dissertation). Victoria University, Australia. Retrieved from https://vuir.vu.edu.au/25850/

Ustabulut, M. Y. (2021). Examination of education program literacy levels of instructors teaching Turkish as a foreign language. *Fırat University Journal of Social Sciences*, *31*(3), 1235-1243. https://doi.org/10.18069/firatsbed.941957

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2005). Practical machine learning tools and techniques. In *Data mining* (2nd ed., Vol. 2, pp. 403-413). Elsevier.

Yıldız, Ü., & Tepeli, Y. (2014). A study on teacher qualifications in teaching Turkish as a foreign language. *International Journal of Language Academy*, *2*(4), 564-578. https://doi.org/10.18033/ijla.182

Article Information Form

Ethical Approval: It is declared that during the preparation process of this study, scientific and ethical principles were followed and all the studies benefited from are stated in the bibliography. The ethics committee report for this study was obtained from the Istanbul University Rectorate Social and Human Sciences Research Ethics Committee with the decision dated 11.01.2024 and numbered 2024/19.

Plagiarism Statement: This article has been scanned by iThenticate.

Appendix 1

Instructor ID	Success Percentage
1	57.5
2	52.5
3	35
4	22.5
5	32.5
6	50
7	30
8	25
9	42.5
10	40
11	32.5
12	50
13	25
14	52.5
15	35
16	35

Appendix 1 (Continued)

17	27.5
18	55
19	25
20	37.5
21	42.5
22	32.5
23	17.5
24	40
25	25
26	55
27	52.5
28	52.5
29	25
30	30
31	42.5
32	72.5
33	40
34	47.5
35	32.5
36	30
37	40
38	37.5
39	50
40	50
41	47.5
42	57.5
43	65
44	42.5
45	375
46	35
47	50
48	475
49	175
50	40
51	22.5

Research Article

Mapping Global Research on Early Childhood Special Education: A Bibliometric Analysis

Gamze İnci [ID]

Kütahya Dumlupınar University, Kütahya, Türkiye, gamzenci@gmail.com

Abstract: Early childhood special education is a scientific field that is gaining worldwide attention. This study uses bibliometric and social network analysis methods to analyze early childhood special education research published in international journals. Therefore, the articles in the WoS database were accessed. WoS is a database of the most prestigious journals indexed SSCI, SCI-Expanded, AHCI, and ESCI. A total of 2051 articles meeting the determined criteria were included in the research. We used the VoSviewer software to perform the analyses and then interpreted the results. According to the research findings, the number of works published in the related field has increased significantly. The Journal of Autism and Developmental Disorders has emerged as the most prolific publisher of articles on this topic. The United States stands out as the most productive country. Dawson, G. is the most effective and cited author. The most frequently used word in the studies examined was autism. It appears that political developments in the field of early childhood special education may contribute to this increase. In addition, the analysis revealed that a certain number of leading countries and researchers have made significant contributions to the field. İnfluential journals in the field of early childhood special education studies, it is noted, focus on special education.

Keywords: Early Childhood, Special Education, Bibliometric Analysis

1. Introduction

Early childhood special education research has an essential place in the scientific literature of the 21st century (Odom & Wolery, 2003). Based on the theoretical foundations of special education and preschool education, this field draws from numerous disciplines, including psychology, social services, child development, sociology, pediatrics, and the health sciences. Therefore, although research on early childhood special education does not have a long history, it has made progress in a short period and gained a prominent position in the scientific literature regarding quantity and quality (McLean et al., 2016; Safford et al., 1994). In recent years, we have observed an increased use of early childhood special education, particularly in policy and scientific research discourses. This shift reflects how many nations adopted the concept of early childhood special education in their social, health, and education policies during the 20th century (McLean et al., 2016; McWilliam, 2016).

The period between the ages of 0 and 6 encompasses early childhood, which is crucial to an individual's development. In addition, the pertinent literature indicates that the age range in early childhood special education practices spans 0 to 8 years (Bowe, 2007; Dunlap, 2005). Early childhood special education, on the other hand, refers to all services provided to children with developmental disabilities or disadvantages at this age and their families. Early childhood intervention focuses on factors influencing an infant's overall biological, social, and academic abilities. The primary objective of early childhood intervention is to provide services that mitigate the adverse effects of the situation faced by young children with special needs who are at risk in terms of development and family stability. In the service delivery procedure, steps are taken to ascertain the situation, evaluate it, and provide suitable services (Oser & Cohen, 2003).

While other countries' legislation presents different classifications, early childhood special education is defined as an umbrella concept that includes early intervention for children aged 0-36 months and preschool special education services for those aged 36-72 months (Dunst & Espe-Sherwindt, 2016). This classification appears in direct practices, especially in the United States (Odom & Wolery, 2003). Two

federal programs were outlined in 2004. Individuals with Disabilities Education Act provides services and support to individuals with special needs. Part B includes regulations concerning formula grants, which aid states in delivering a free, adequate public education in the least restrictive environment for children with impairments aged three to 21. Part C of the Individuals with Disabilities Education Act covers services for children birth to three years of age (2004). Providers commonly deliver these services, known as Early Intervention (EI), in homes or childcare settings, focusing on a family-centered approach (Bruder, 2016). The primary objective of EI services is to equip families with intervention and development support skills for children with disabilities (McWilliam, 2016). The three-year-old transition from Part C to Part B services (Individuals with Disabilities Education Act 2004). We commonly refer to this phase as Early Childhood Special Education (ECSE), and programming shifts from a family- and home-centered to a child- and school-centered system (Bruder, 2016).

Policymakers have also been affected by the shift in perspective and approach toward individuals with special needs, particularly during the middle of the 20th century. This has led to the establishment of crucial legal regulations. Examining the scientific studies reveals that there has been a significant increase in scientific studies since the enactment of legislation in the relevant field (Odom & Wolery, 2003). When analyzing the historical development of early childhood special education, the political, economic, and scientific conditions of the period during the developmental stages are highlighted.

Although early childhood special education has an interdisciplinary feature, it also has a dynamic structure that constantly renews itself (Guralnick, 2005). Due to this, it represents a significant advancement in scientific writing at the end of the 21st century. Today, both political and scientific interest in the field is burgeoning. Therefore, it takes more and more time and effort for researchers to access relevant studies in this field. Determining what kind of progress has been made in the literature and what scientific trends are can help provide readers and stakeholders with perspective and insight into the field's developments. Although early childhood special education has an interdisciplinary feature, it also has a dynamic structure that constantly renews itself (Guralnick, 2005). Due to this, it represents a significant advancement in scientific writing at the end of the 21st century. Today, both political and scientific interest in the field is burgeoning. Therefore, it takes more and more time and effort for researchers to access relevant studies in this field. Evaluating the advancements achieved by the literature and what scientific trends are can help provide readers and stakeholders with perspective and insight into the field's developments. From this perspective, it can be said that the current study will provide a significant competitive advantage to researchers, institutions, and countries by accessing accurate information promptly. Recently, interest in early childhood special education research has grown exponentially. As statistical findings have accumulated, academics have analyzed and summarized the published papers to give researchers, educators, and policymakers an overview of the field's present status and developments. To date, researchers have mostly explored this topic through literature reviews and content analysis of scholarly publications. These analyses examine and characterize the conceptual frameworks, theoretical perspectives, and paradigms underlying early childhood special education literature. Numerous authors interested in the field have provided extensive reviews in various ways, such as meta-analysis and systematic review, in studies that conceptualize early childhood special education areas, discuss their different dimensions, and attempt to identify effective practices (e.g., Banerjee et al., 2018; Bonuck et al., 2011; Gül-Olcay & Diken, 2009; Öncül, 2014; Sender & Sheehan, 1983; Yıldırım & Akçamete, 2014; Virués-Ortega, 2010). Such systematic literature reviews, meta-analyses, meta-synthesis studies, and content analysis compilations have undoubtedly significantly contributed to the field. However, due to the nature of these methods, studies involving them necessitate substantial investments of time and resources, resulting in a limited number of publications to be reviewed. In addition, the interpretation of the studied information is highly subjective, limiting the generalizability of these studies' findings. Bibliometric studies offer an alternative to systematic literature research and content analysis by increasing the number of studies

that researchers can examine while reducing the time and effort needed for data processing (Hernández-Torrano, 2020). Bibliometric studies employ coded bibliographic data from databases to map and identify a study topic's current state and evolution by examining publication and citation trends. In addition, it evaluates efficiency rankings for journals, countries, institutions, and authors, as well as the detection of similarity and collaboration trends among academics (Andrés, 2009).

Despite the growing interest in early childhood special education research, no study has used bibliometric methods to summarize the available literature. This study aims to fill this gap by providing a comprehensive overview of the current state and development of research on special education for young children over the past 46 years. In this context, the primary objective of this study is to present the findings of a bibliometric analysis of early childhood special education articles published in the Web of Science (WoS) database between 1982 and the present (2021, March). The current study aims to identify trends among active researchers, countries, institutions, and groups by revealing the bibliometric characteristics and social network structure of early childhood special education studies. As a result, in this study, we will identify the frequently studied topics, prolific authors, countries, institutions, and primary reference sources in early childhood special education.

2. Method

Using metadata from the Web of Science (WoS) database, this study aimed to analyze and map early childhood special education literature over the past 46 years using bibliometric methods. Bibliometric analysis is a quantitative method for analyzing the informational content of large data sets (White & McCain, 1998; Zupic & Cater, 2015). Although this analysis method is widely utilized in other fields, it has only recently begun to be implemented in education (e.g., Hallinger & Kovaevi, 2019; Hernández-Torrano, 2020). Moreover, when specifically investigating the field of special education, we find that studies employing this methodology are scarce.

This study section will describe data sets and data analysis methods. Due to the readership's unfamiliarity with the bibliometric analysis method, the subheadings provide detailed information on the creation of data sets, the data analysis process, and the mapping and synthesis of information.

2.1. Creation of the database

In accordance with the study's primary purpose, we conducted a systematic search in the WoS Core Collection database, which contains high-quality and scientific research in more than 250 scientific, social, and humanities disciplines. The analysis incorporated the Social Sciences Citation Index (SSCI), the Arts & Humanities Citation Index (A&HCI), the Emerging Sources Citation Index (ESCI), and the Science Citation Index- Expanded (SCI-Expanded), which are the four fundamental indexes in the cited database.

Bibliometric analyses and science mapping are based on the analysis of scientific content extracted from digital databases like WoS, Scopus, and Google Scholar (Halinger, 2020). In this study, we selected the WoS database for several essential reasons. First, a well-known fact that WoS is the most published and

authoritative database used in bibliometric analysis (Grosseck, Tîru, & Bran, 2019; Meho & Yang, 2007; Mongeon & Paul-Hus, 2015). In addition, WoS has multidisciplinary content that presents data related to early childhood special education in more than 20,000 different disciplines and research areas. Compared to other databases, WoS data includes citation information for studies published in broader historical contexts (Liu et al., 2019).

In the systematic review processes for providing data documents, we adhered to the "Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)" (Moher et al., 2009) recommendations. Figure 1 represents using the PRISMA methodology. In addition, the present study

followed the steps suggested by Zupic and Cater (2015) for mapping studies. These five stages are research design, bibliometric data collection, data analysis, data visualization, and interpretation.

Initial research in the WoS database for data collection focused on two main areas. The first of these we use in the title, and the second in the topic area. We conducted an advanced search to identify targeted studies in the two areas. Such advanced scanning options in the WoS database allow for criteria specification and combination creation. In this study, researcher used the keywords "Early Childhood Special Education," "Early Intervention," "Young Children," and "Early Childhood Intervention" in the title area for scanning. They used the keywords "Special Education," "Inclusion," "Mental Retardation," "Autism," "Disability," and "Down Syndrome" in the topic area. This related dataset obtained was combined using AND. We specified no date range in the review. Thus, we aimed to reach all the studies indexed in the WoS database, starting from the oldest article in this field until the search date of March 30, 2021. As a document type, we included only article studies in the search. We did not limit any research area and included studies in all areas in the data collection process. The inclusion criteria used in the study are presented in Table 1.

Table 1

Criteria for the Selection of the Publications

Criteria	Value
Data Source	Web of Science
Search Terms	"Early Childhood Special Education" OR "Early Intervention" OR "Young Children" OR "Early Childhood Intervention" AND "Special Education" OR "Inclusion" OR "Mental Reterdation" OR "Autism" OR "Disability" OR "Down Syndrome"
Country	All Country
Document Type	Article
Citation Index	SSCI, ESCI, A&HCI, SCI-Expanded
Language	All Language
Categories	All WoS Categories

We reached a total of 2,051 articles that met the above inclusion and exclusion criteria. Figure 1 shows the flow of the dataset creation process.

Figure 1

Flow Diagram of Determination and Selection of Publications
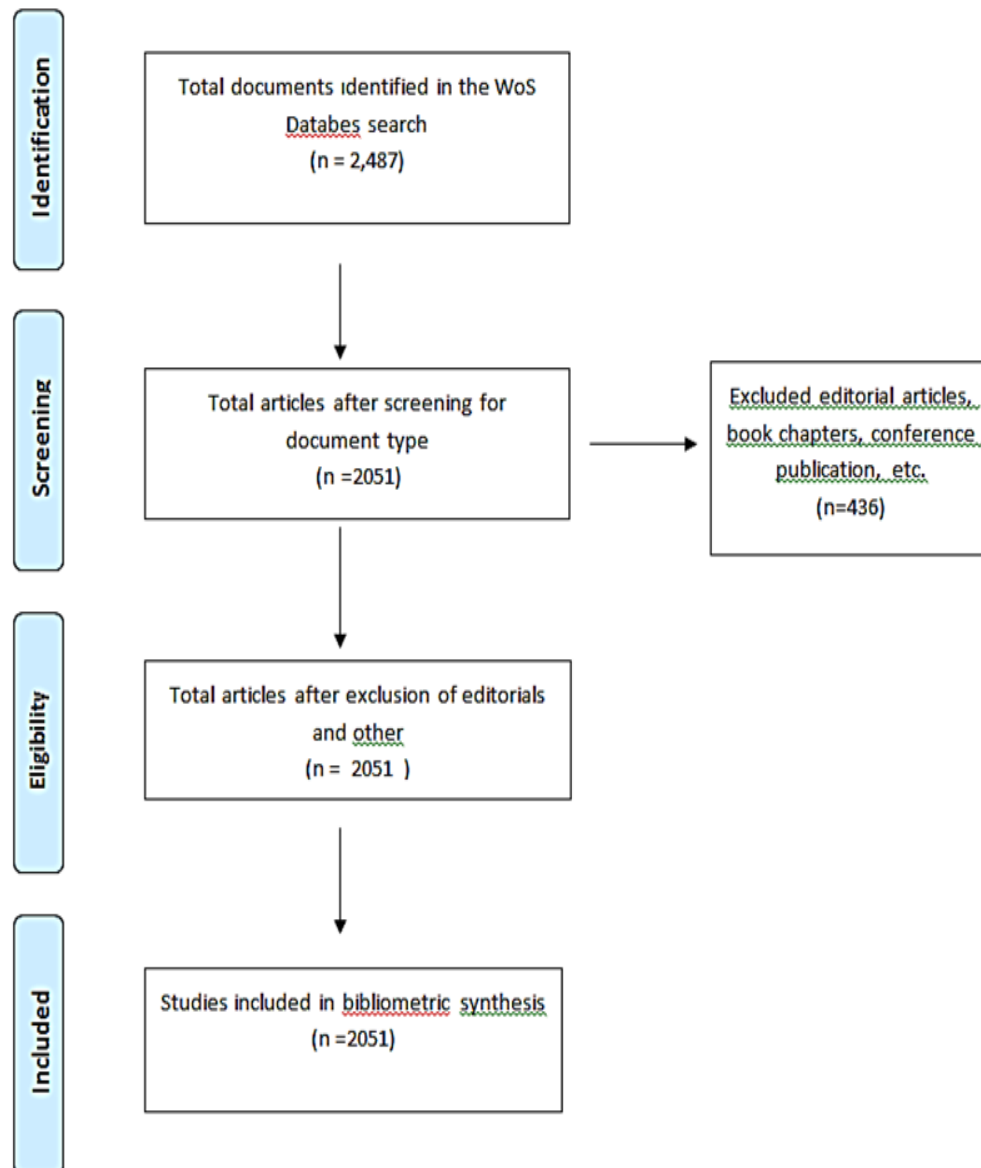
2.2. Data analysis

Science mapping employs descriptive statistics, citations, and social network analysis to comprehend a study field's evolution, composition, and intellectual makeup (Zupic & Cater, 2015). In this study, we analyzed descriptive and bibliometric data to offer an overview of the evolution and current status of scientific research in early childhood special education research from the past to the present.

To reveal the evolution and expansion of the subject, we used descriptive statistics to identify the distribution of the number of articles and citations by year, as well as the frequency of citations and publications by the most productive journals, authors, and nations. For this analysis, we conducted the citation analysis. Citation analysis measures the relative importance or impact of an author, an article, or a publication by counting the number of times other works have cited that author or publication. Within the scope of our study topics, we also incorporated the social network analysis method, which is commonly utilized in bibliometric approaches. We used social network analysis techniques to disclose

the word network structure, journal co-authorship analysis network structure, source co-authorships analysis network structure, and author co-authorships analysis network structure in the articles analyzed for the research topics to illustrate the data. Research in social network analysis is becoming increasingly valuable to the social sciences, particularly sociology (Otte & Rousseau, 2002). This method can also explain unknown conditions in the associated area, which is one of the reasons for the increased interest in network research in numerous fields (Kilduff & Tsai, 2008). Approaches to bibliometric network analysis employ diverse ways of analysis (e.g., Co-authorship analysis, Co-citation analysis, Bibliographic coupling, etc.). Citation analysis, Co-author analysis, and co-occurrence were utilized in the current study to obtain relevant solutions to the research questions.

Co-author analysis examines the social structure of a subject or research topic. In co-authorship analysis, networks at the level of institutions and countries emerge. Co-author analysis reveals cooperation in the area (Zupic & Cater, 2015) and displays the intellectual organization of a field (Small, 1973). In the current study, co-authorship network analyses were undertaken to assess the collaboration between authors, institutions, and nations in early childhood special education scientific journals.

Finally, we undertook a co-occurrence analysis of terms to highlight the conceptual framework of early childhood special education. Keyword co-occurrence analysis is a form of content analysis that examines the conceptual structure of a research field (Collan et al., 1983). Co-occurrence analysis of keywords provides insight into the most researched subjects and concepts (Zupic & Čater, 2015).

We evaluated the data using version 1.6.7 of VOSViewer. VOSViewer is software for creating and visualizing bibliometric networks. This software also leverages text mining algorithms to construct and visually display networks of important phrases extracted from datasets (Van Eck and Waltman 2014). In three steps, the software generates bibliometric maps using a distance-based method (Van Eck & Waltman, 2017). First, the software normalizes the disparities across nodes. The second stage produces a two-dimensional map in which the distance between nodes is proportional to their similarity. In the third step, VOSViewer clusters tightly linked nodes (Van Eck & Waltman, 2014).

3. Findings

This section presents the findings of bibliometric and network analyses to answer the research questions. First, the number of articles and citations of early childhood special education studies is provided. Then, the leading journals in the field and the most cited articles are listed. Third, the authors, institutions, and countries leading the field are presented. Finally, analyses of author collaboration, institutions, countries, and the network of conceptual change in the work from the past to the present are described.

3.1. Publication and citation trends

The number of publications and citations correctly measures the growth and development of a scientific field. The data set derived from early childhood special education studies reveals that 2051 papers were published between 1982 and March 2021. The data set derived from early childhood special education research indicates that 2051 papers were published between 1982 and March 2022. From 1982 to March 2021, Figure 2 demonstrates the dynamic growth trajectory of scholarly articles in early childhood special education. In the 39 years after the initial publication in 1982, there has been a progressive and consistent increase, as seen in Figure 2. The years can be divided into three parts: a period of emergence (1982–1992), during which the number of publications rose slowly; a period of development characterized by a remarkable increase (1993–2005); and a period of growth, during which the number of studies published in the relevant field increases annually (2006-2020). These periods represent typical phases of a new field of study. 2020 saw the greatest number of publications, with 181 articles (Eck & Waltman, 2014).

Figure 2

Number of Publications by Year
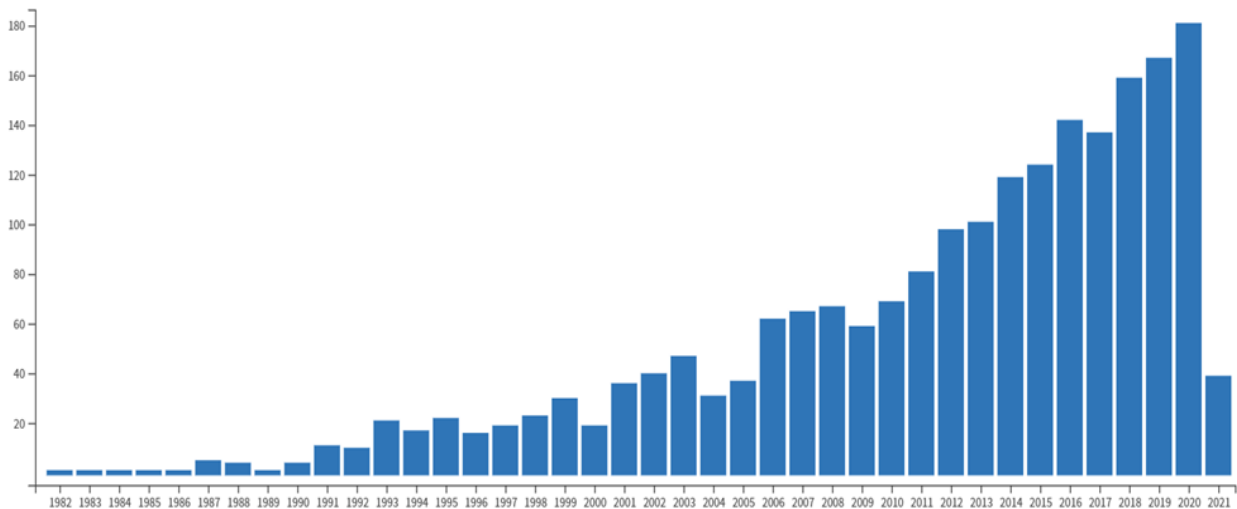
Similarly, the number of citations to research on early childhood special education has increased annually. All associated articles garnered 54,886 citations in total. While the average number of citations per article is about 26,5, the average number of citations per year is 1,444.37. Figure 3 depicts the distribution of citations for relevant publications by year. The number of citations increased somewhat at precise intervals from 1984, when the first citation was made, through 1996, as depicted in Figure 3. Since 1996, there has been a notable increase in citations. Since 2006, the quantity of citations has climbed gradually. The year with the highest number of citations was 2020, with a total of 6512. The growth in publications and citations indicates congruence.

Figure 3

Number of citations by year
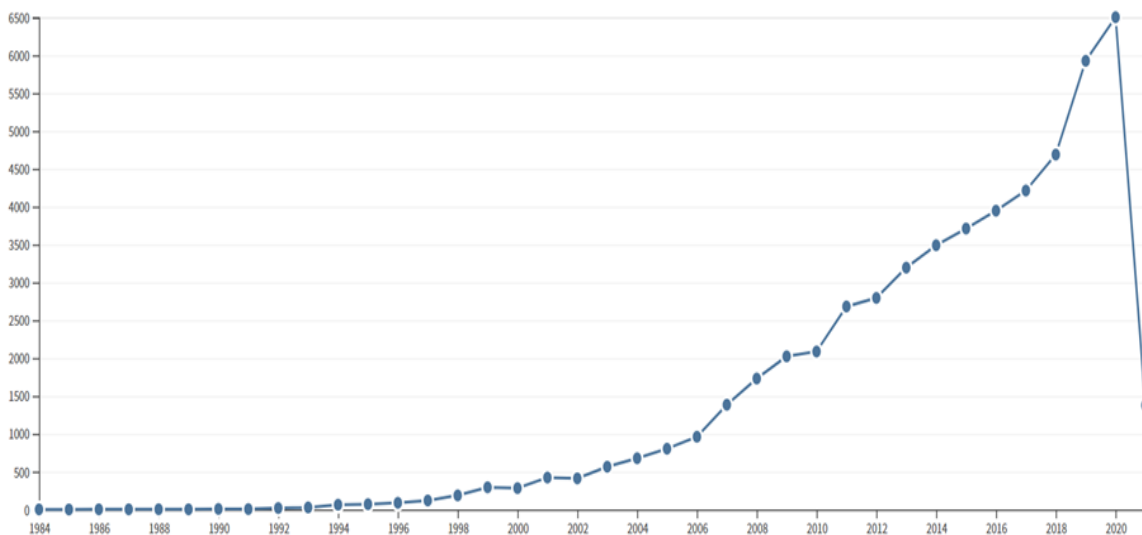
3.2. Core journal and publications

The reviewed 2051 papers were published in 655 journals. Given the number of journals that publish articles, it is essential to emphasize that early childhood special education research is generally

recognized by journals in various fields. Additionally, only 36 of 665 journals have published ten or more relevant publications. According to this perspective, most early childhood special education studies are published in journals linked to the topic. Table 2 displays the ten journals that publish the most research on early childhood special education. Also included in Table 1 are the total number of citations and the connection strengths. These are the major journals in the field of early childhood special education.

Table 2

Most Highly-Cited Journal on Early Childhood Special Education on WoS Citation Analysis, 1982–2021 (n = 655)

Rank	Journal	Number of Documents	WoS Cites	Total Link Strength
1	"Journal of Autism and Developmental Disorders"	160	8192	562
2	"Topics in Early Childhood Special Education"	81	1840	218
3	"Infants & Young Children"	74	1251	156
4	"Journal of Early Intervention"	67	1327	199
5	"Autism"	63	1542	238
6	"Research in Autism Spectrum Disorders"	63	1269	191
7	"Research in Developmental Disabilities"	47	1574	145
8	"Journal of Developmental and Physical Disabilities"	28	472	44
9	"Autism Research"	26	386	67
10	"Journal of Intellectual Disability Research"	23	1077	58

With 160 articles, the "Journal of Autism and Developmental Disorders" has the most amount of early childhood special education publications. In addition, the same journal has the maximum number of citations (8192) and link strength (727) relative to the number of publications. "Topics in Early Childhood Special Education", with 81 articles, and "Infants & Young Children", with 74 articles, are the two journals that rank second and third, respectively, in terms of the number of articles published. The remaining articles in the collection are dispersed throughout numerous periodicals in diverse fields, including health, psychology, and sociology.

Table 3 lists influential publications in the early childhood special education discipline based on the number of citations. This table displays the top ten most cited publications in the data set obtained from the WoS database. Between 2000 and 2006, nine journals published the ten most influential articles. The total number of citations to these publications is 4564, representing 8.3% of all citations in the data set. "The Journal of Child Psychology and Psychiatry" has published two of the most-cited papers.

Table 3

Most Highly-Cited Early Childhood Special Education Publications on WoS Citation Analysis, 1982–2021.

Rank	Title	Authors	Journal	Year	Citations
1	"Identifying infants and young children with developmental disorders in the medical home: An algorithm for developmental surveillance and screening"	"Duby et al."	"Pediatrics"	2006	772
2	"Brain structural abnormalities in young children with autism spectrum disorder"	"Sparks et al."	"Neurology"	2002	581
3	"Changing illness perceptions after myocardial infarction: An early intervention randomized controlled trial"	"Petrie et al."	"Psychosomatic Medicine"	2002	540
4	"Long-term effects of an early childhood intervention on educational achievement and juvenile arrest - A 15-year follow-up of low-income children in public schools"	"Reynolds et al."	"Journal Of the American Medical Association"	2001	433
5	"Joint attention and symbolic play in young children with autism: a randomized controlled intervention study"	"Kasari et al."	"Journal of Child Psychology and Psychiatry"	2006	429
6	"Sensory Experiences Questionnaire: discriminating sensory features in young children with autism, developmental delays, and typical development"	"Baraneck et al."	"Journal of Child Psychology and Psychiatry"	2006	404
7	"Randomized trial of intensive early intervention for children with pervasive developmental disorder"	"Smith et al."	"American Journal on Mental Retardation"	2000	404
8	"The behavioral phenotype in fragile X: Symptoms of autism in very young children with fragile X syndrome, idiopathic autism, and other developmental disorders"	"Roger et al."	"Journal of Developmental and Behavioral Pediatrics"	2001	348
9	"Age and IQ at intake as predictors of placement for young children with autism: A four- to six-year follow-up"	"Harris and Handleman."	"Journal of Autism and Developmental Disorders"	2000	330
10	"Neural correlates of face and object recognition in young children with autism spectrum disorder, developmental delay, and typical development"	"Dawson et al."	"Child Development"	2002	323

3.3. Leading authors and countries

The data set contains 2051 articles authored by 6917 people from 101 countries and 1973 institutions worldwide. Table 4 shows the ten authors with the earliest childhood special education publications. Dawson G. is the leader with 28 publications. All 28 studies have a total of 2,768 citations. Roger S. J. follows this author with 27 papers and Matson J. L. with 19 articles. Notable is the fact that fifty percent of the authors are from the same country (USA). Given the list's geographical information, it is evident that South America, Africa, and Asia are not included. Lastly, as demonstrated in Table 3, the number of publications and citations do not correspond numerically. Matson J.L. has 430 citations despite having the most papers, whereas Lord C. gets 933 citations despite having just 14 articles.

Table 4

Top Most Prolific Authors Ranked by Total Number of Publications

Rank	Author	Country	Number Of Publications	Number Of Citations
1	Dawson, G.	USA	28	2768
2	Rogers, S.J.	USA	27	2444
3	Matson, J.l.	USA	19	430
4	Barton, E.	USA	15	248
5	Bruder, M.	USA	15	457
6	Mcconachie, H.	England	15	507
7	Gillberg, C.	Sweden	14	345
8	Lord, C.	USA	14	930
9	Ingersoll, B.	USA	13	692
10	Keen, D.	Australia	13	194

2051 documents in which published early childhood special education studies are sourced from a total of 101 countries worldwide. Table 5 shows the ranking of the top ten countries according to the number of publications. As well as the most prolific author ranking, the United States is the most productive country, leading early childhood special education research with 1,281 articles. In the United States, articles produced for the relevant discipline account for 62.422% of all publications. There has been a total of 38186 citations for these studies. This represents more than fifty percent of all article citations. This makes the United States the leader in early childhood special education. Australia, with 200 studies, and England, with 147 research, come in second and third, respectively, after the United States. The top four countries in the ranking all speak English. In addition, the table includes OECD members, including Austria, Canada, Germany, and the Netherlands.

Table 5

Top Countries by Number of Publications in the Dataset

Rank	Country	Documents	Citations
1	USA	1281	38186
2	Australia	200	4.057
3	England	147	5.076
4	Canada	108	2945
5	Sweden	55	1879
6	Netherlands	52	912
7	Italy	44	1266
8	Germany	37	791
9	People S R China	36	340
10	Israel	36	1158

3.4. Research collaborations between authors, institutions, and countries (Co-authors analysis)

In this section, we utilized co-authorship analysis to discover scientific collaboration networks between scholars, institutions, and countries in early childhood special education. Figures 4, 5, and 6 depict author, institution, and country co-authorship networks in early childhood special education.

The nodes shown in the social network analysis reflect the units of analysis (authors, institutions, and countries). The number of broadcasts corresponds to the size of the nodes. The relationship between nodes is defined by their distance apart. When the distance between two nodes decreases, a strong link is deemed to exist. The color of the nodes is used to distinguish across clusters and denotes which cluster the node belongs to. The thickness of the networks that comprise the connections is indicative of the relationship's strength (Van Eck & Waltman, 2014).

Figure 4 shows collaborative networks among authors with at least five published works (n=94). This map revealed a network of five or more researchers who contributed to the advancement of the field. The map shows a total of 36 clusters. Authors in the red, blue, and green clusters established the most extensive collaboration network. These clusters of seven scholars each represent an international research group. Other massive clusters are yellow and purple, each containing six researchers. These clusters serve as the point of connection for numerous other researchers.

Figure 4
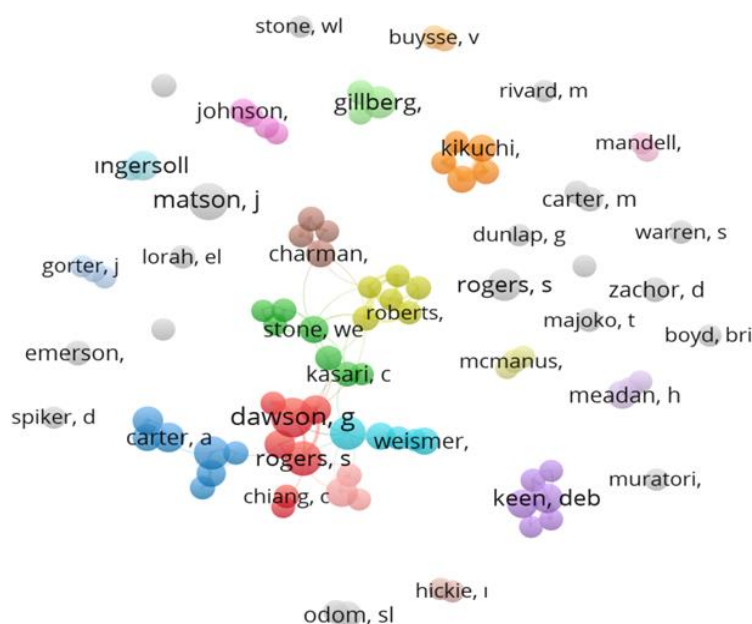
Map of Scientific Collaborations Between Researchers

Figure 5 maps scientific networks among institutions that have published 15 or more articles in the dataset (n=42). The map is divided into 6 clusters. Institutional cooperation networks are generally formed within national borders and geographically close institutions. Clusters in the map's central horizontal region link institutions in the United States and Canada (red, blue, green). Inter-cluster connections are also seen high in the mentioned areas. Interestingly, despite its remote location, Tel Aviv University appears to collaborate with institutions in the United States. Additionally, while several links arise between institutions within the same cluster, the density of linkages between institutions from other clusters drops.

Figure 5
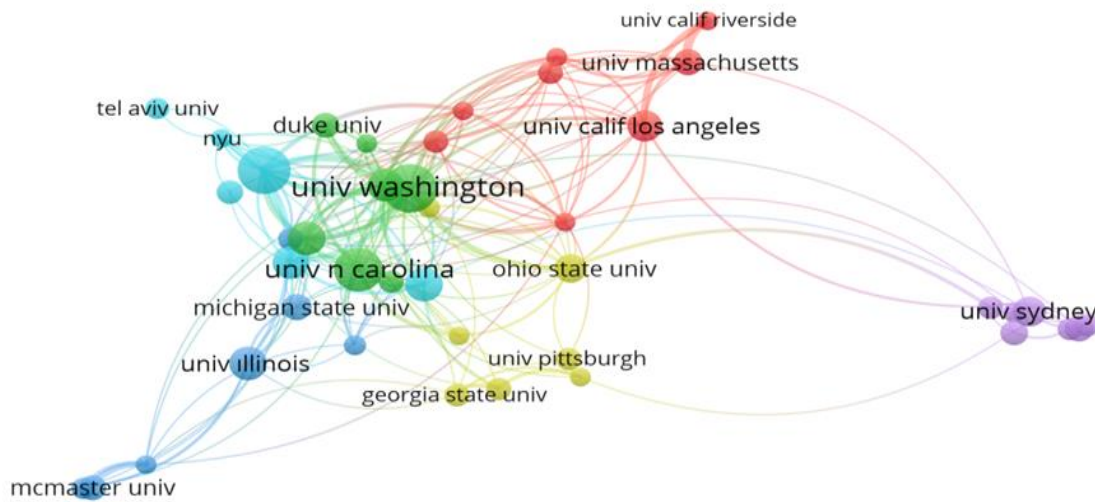
Map of Scientific Collaborations Between Institutions

Figure 6 shows a map of international cooperation in early childhood special education research. This map examines research collaborations (n=35) between countries with at least five publications. The United States represents the largest connecting node in the geographic center. Additionally, it shares cooperative ties with all other countries/regions, forming a cluster with Greece, South Korea, Turkey, and Taiwan. Although there are six clusters on the map, it can be said that international collaborations are generally shaped by geographical and linguistic proximity. For instance, the red cluster has the greatest number of countries (13) and consists of European nations, including France, Germany, Italy, Denmark, Belgium, and the Netherlands. Globally, the data indicate that even though there is some cooperation across different regions of the world, international collaboration networks are still determined by geographic closeness.

Figure 6

*Map of scientific collaborations between countries*
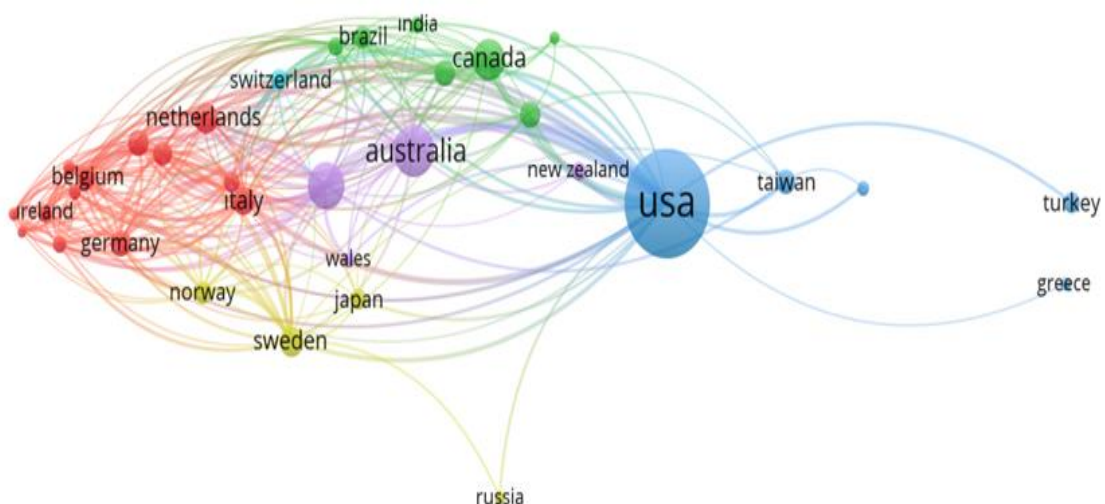
3.5. Topical foci of research on early childhood special education

Figure 7 illustrates the network analysis of the present research that focuses on all article studies in the early childhood special education field between 1982 and March 2021. This study helps us develop a conceptual map by identifying the primary topics of interest to scholars in the relevant subject. The size of the nodes on the map represents the frequency of author keywords within the dataset, while the thickness of the borders represents the co-occurrence strength of keyword pairs. We included seventy-nine words with at least ten co-occurrences in the analysis. The analysis resulted in the formation of seven clusters. Autism is the most widely used keyword, with 390 occurrences. It is followed by "early intervention" formations numbering 93.

Figure 7

Map of topical foci of research on early childhood special education
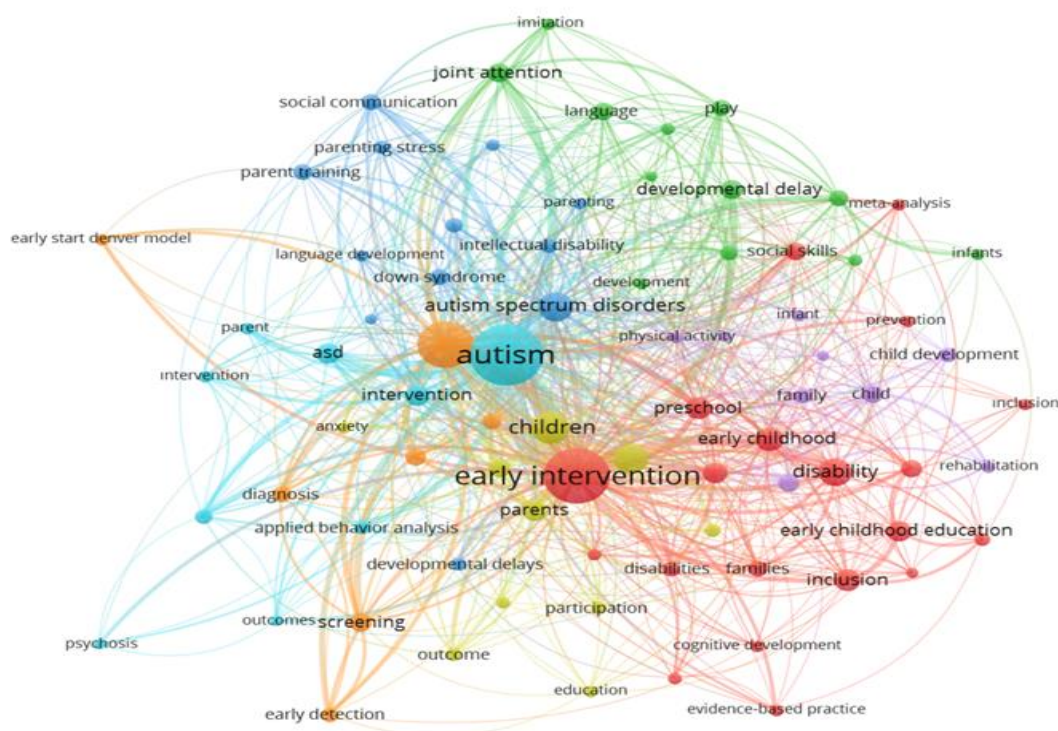
4. Discussion and Conclusion

Publications and citations have increased progressively during the past 39 years, and their exponential growth over the past 15 years reflects the anticipated growth trajectory of a new field of study. Although the first study in early childhood special education was published in 1982, the number of studies in the field has expanded dramatically since 1992. This growth is proportional to the number of articles and citations. On the other hand, the influence of political developments in the field on scientific literature is evident. As a matter of fact, the 1990s marked the beginning of the legalization of early childhood special education (Bricker, Xie, & Bohjanen, 2018). These laws, which began in 1990 and continue to this day, are considered a turning point in early childhood special education theory and practice.

Another important factor contributing to this dramatic rise in publications and citations is an increase in journals published related to early childhood special education. Today, many international journals aim to publish research on early childhood interventions and special education. In addition, easy access to information and extensive usage of internet databases contribute to the quantitative expansion of the

discipline. Given this information, it is anticipated that the number of early childhood special education studies will continue to rise in the years to come.

According to the research data, the most productive journal in the field is the "Journal of Autism and Developmental Disorders," with a number of publications and citations. Although the journal is not directly focused on early childhood special education, the results of the analysis show that it significantly impacts the field. Similarly, four journals focusing on autism studies are among the ten most productive early childhood special education research journals. This shows that early childhood studies have an essential place in autism research. Indeed, the importance of early intervention is emphasized in autism research (Koegel et al., 2014; Zwaigenbaum et al., 2013; Webb & Jones, 2009). We can show this situation as one of the crucial reasons early childhood comes to the fore in studies on autism. "Topics in Early Childhood Special Education," "Infant Young Children," and "Journal of Early Intervention" are other journals that publish the most in the field of early childhood special education. The fact that these journals have a framework that directly acquires the field of early childhood special education is considered an essential effect. Even though early childhood special education is a frequent subject of study in various disciplines, the top 10 journals with the most significant number of articles in the field serve a variety of special education subfields.

It is interesting that there is no correlation between those mentioned above and the ten most prominent journals and research based on the number of citations related to early childhood special education. A publication that appeared exclusively in the "Journal of Autism and Developmental Disorders" garnered 330 citations and joined the top ten most-cited papers. Examining the publications reveals that the top nine articles with the most citations have several authors, whereas only one has two. However, it is noteworthy that more than half of the publications are in medical journals. Similarly, we may state that the most cited papers in the early childhood special education field contain information on medical diagnosis and evaluation. This is an important result of the early childhood special education field's complex structure and interdisciplinary nature (McCormick & Brennan, 2001). Aboelela et al., (2007) have suggested that multidisciplinary research is a practical approach to solving complex problems, providing critical answers, and facilitating the application of knowledge in a particular field.

Considering the number of publications and citations, examining the most prolific researchers in the field of early childhood special education reveals that some authors have made substantial contributions to the field. Dawson, G., Rogers, SJ., and Matson, Jl. It has almost half the total number of publications among the top ten authors in the field. Similar findings have been found in bibliometric studies conducted in various disciplines (Cretu & Morandau, 2002). The process of accumulative advantage can explain variations in publication and productivity distributions across researchers in the same field. This technique boosts the output of influential scientists while decreasing the productivity of low-producing scientists due to resource recognition and feedback. Cross-sectional study data demonstrate that increasing career age increases academic production inequality (Allison & Stewart, 1974; Wang et al., 2006).

Early childhood special education is an international topic of study that has sparked political and scientific agendas across different eras and regions of the world, driven by the needs and experiences of local populations (Smith, 2000). Examining the research findings to support this prediction reveals that different countries contribute to the field in question. In addition, research indicates that the United States is the most productive nation in producing information and researching this subject. This finding is bolstered by the fact that seven of the ten most prolific researchers on the subject are American. Three English-speaking nations follow the United States. These are Australia, the UK, and Canada, respectively. One important reason for this finding could be that the searched database only contains English sources. The results show that many Asian and African countries are far behind in productivity. Reasons such as

language barriers, insufficient financial resources for scientific studies, and limited access to data may be the reason (Wang et al., 2006).

There is an increasing research interest based on analyzing research collaboration (Kim, 2006). Even if it is not a direct indicator of quality, research collaboration is an essential and valid method for achieving it. The analysis of co-authorship data in this study revealed the network structure of field collaboration in relation to nations, institutions, and researchers. Research collaboration is viewed as a sign of quality research or a way of reaching quality and a tool to stimulate and support research in places with less developed research infrastructure and capabilities (Katz & Martin, 1997; Kim, 2006).

When we generally examine the findings from this study, some collaborative research networks in early childhood special education research at the researcher, country, and institution level do not seem well-established. According to the analyses, researchers typically work alone or in small groups. Examining international collaboration reveals that the United States is at the center. This circumstance demonstrates a correlation between the number of publications and citations. Additionally, geographical and cultural proximity is another factor that influences cooperation (Zitt et al., 2000).

Examining the social network structure for institutional collaboration reveals that the studies are conducted within national borders. This analysis demonstrates that the networks connecting geographically close institutions are robust. This implies that researchers collaborate with other researchers in their home countries on an individual and institutional level. This network structure is observed in numerous social science disciplines (Mosbah-Natanson & Gingras, 2014).

Keyword analysis gives a conceptual map of research on early childhood special education. Since early childhood special education is an interdisciplinary field of study, many notions have a complicated, variable-size network structure. Autism spectrum disorder, in particular, has the most extensive social network. Additionally, keywords such as parent education, evidence-based practices, developmental delay, language acquisition, and early intervention are commonly used with research in this field.

In conclusion, the present study mapped the existing literature in the early childhood special education field over the past 39 years. According to the results obtained from the data, the field has been growing exponentially, especially in the last ten years. All the same, it is seen that this field is a dynamic scientific discipline that is constantly developing. Although some geographic locations stand out, early childhood special education research is a field that draws global attention. Additionally, improved cooperation between countries, scholars, and institutions in the subject may substantially affect the field's advancement and application.

5. Limitations

Even though bibliometric studies provide a basic overview of the study subject, they cannot capture all the field's particulars. We conducted this investigation within the constraints of certain restrictions. The first constraint is that we limited the search to the WoS database alone. The limitations of the WoS database may also apply to this research. As a result, future analyses should rely on data from numerous sources, such as Scopus. Researchers can expand future studies by using a database in Which different disciplines are scanned, considering the field's interdisciplinary nature. Another limitation is the inclusion of only article-type publications in the current study. The research excluded books, book chapters, conference papers, and other documents that have contributed to the field of early childhood special education. An additional study can be conducted to compile a summary of various scientific documents. Also, in this research, bibliometric data covers articles published until March 2021. Therefore, it can be said that it does not fully reflect 2021. In addition, in analyzing bibliometric data, citation-based indicators also consider the author's or author group's self-citations. The authors' locations, language, and self-citations can all affect citations. All this makes bibliometric studies sometimes biased and manipulable. Funding: No funding source is reported for this study.

References

Aboelela, S. W., Larson, E., Bakken, S., Carrasquillo, O., Formicola, A., Glied, S. A., ... & Gebbie, K. M. (2007). Defining interdisciplinary research: Conclusions from a critical review of the literature. *Health services research*, *42*(1p1), 329-346. https://doi.org/10.1111/j.1475-6773.2006.00621.x

Allison, P., & Stewart, J. (1974). Productivity differences among scientists: Evidence for accumulative advantage. *American Sociological Review, 39*(4), 596-606. https://doi.org/10.2307/2094424

Andrés, A. (2009). *Measuring academic research: How to undertake a bibliometric study.* Elsevier.

Banerjee, R., Movahedazarhouligh, S., Millen, K., & Luckner, J. L. (2018). Building evidence in early childhood special education: A systematic review of replication intervention studies. *Topics in Early Childhood Special Education, 37*(4), 246-256. https://doi.org/10.1177/0271121417748197

Bonuck, K. A., Hyden, C., Ury, G., Barnett, J., Ashkinaze, H., & Briggs, R. D. (2011). Screening for sleep problems in early intervention and early childhood special education: A systematic review of screening and assessment instruments. *Infants & Young Children, 24*(4), 295-308. https://doi.org/doi:%2010.1097/IYC.0b013e3182234444

Bowe, F. G. (2007). *Early childhood special education: Birth to eight* (4th ed.). Thomson Delmar Learning.

Bricker, D., Xie, H., & Bohjanen, S. (2018). A history of EI/ECSE in the United States: A personal perspective. *Journal of Early Intervention, 40*(2), 121-137. https://doi.org/10.1177/1053815118771392.

Bruder, M. B. (2016). Personnel development practices in early childhood intervention. In B. Reichow, B. A. Boyd, E. E. Barton, & S. L. Odom (Eds.), *Handbook of early childhood special education* (pp. 289-333). Springer International Publishing.

Callon, M., Courtial, J. P., Turner, W. A., & Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information, 22*(2), 191-235. https://doi.org/10.1177/053901883022002003

Cretu, D. M., & Morandau, F. (2020). Initial teacher education for inclusive education: A bibliometric analysis of educational research. *Sustainability, 12*(12), 4923. https://doi.org/10.3390/su12124923

Dunlap, L. L. (2005). *An introduction to early childhood special education: Birth to age five.* Pearson Education.

Dunst, C. J., & Espe-Sherwindt, M. (2016). Family-centered practices in early childhood intervention. In B. Reichow, B. A. Boyd, E. E. Barton, & S. L. Odom (Eds.), *Handbook of early childhood special education* (pp. 37-56). Springer International Publishing.

Grosseck, G., Țîru, L. G., & Bran, R. A. (2019). Education for sustainable development: Evolution and perspectives: A bibliometric review of research, 1992–2018. *Sustainability, 11*(21), 6136. https://doi.org/10.3390/su11216136

Guralnick, M. J. (2005). *The developmental systems approach to early intervention.* Brookes Publishing.

Gül-Olcay, S., & Diken, İ. H. (2009). A review of master and doctoral dissertations in the field of early childhood special education in Turkey. *International Journal of Early Childhood Special Education, 1*(1), 46-78. https://doi.org/10.20489/intjecse.107977

Hallinger, P. (2020). Science mapping the knowledge base on educational leadership and management from the emerging regions of Asia, Africa, and Latin America, 1965– 2018. *Educational Management Administration & Leadership, 48(*2), 209-230. https://doi.org/10.1177/1741143218822772

Hallinger, P., & Kovačević, J. (2019). A bibliometric review of research on educational administration: Science mapping the literature, 1960 to 2018. *Review of Educational Research, 89*(3), 335-369. https://doi.org/10.3102/0034654319830380

Hernández-Torrano, D. (2020). Mapping global research on child well-being in school contexts: A bibliometric and network analysis (1978–2018). *Child Indicators Research, 13*(3), 863-884. https://doi.org/10.1007/s12187-019-09659-x

Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy, 26*(1), 1-18. https://doi.org/10.1016/S0048-7333(96)00917-1

Kilduff, M. & Tsai, W. (2008). *Social networks and organizations.* Sage.

Kim, K. W. (2006). Measuring international research collaboration of peripheral countries: Taking the context into consideration. *Scientometrics, 66*(2), 231–240. https://doi.org/10.1007/s11192-006-0017-0

Koegel, L. K., Koegel, R. L., Ashbaugh, K., & Bradshaw, J. (2014). The importance of early identification and intervention for children with or at risk for autism spectrum disorders. *International Journal of Speech-language Pathology, 16*(1), 50-56. https://doi.org/10.3109/17549507.2013.861511

Liu, W. (2019). The data source of this study is Web of Science Core Collection? Not enough. *Scientometrics, 121*(3), 1815-1824. https://doi.org/10.1007/s11192-019-03238-1

McCormick, K. M., & Brennan, S. (2001). Mentoring the new professional in interdisciplinary early childhood education: The Kentucky Teacher Internship Program. *Topics in Early Childhood Special Education, 21*(3), 131-149. https://doi.org/10.1177/027112140102100301

McLean, M., Sandall, S.R., & Smith, B.J. (2016). A history of early childhood special education. In B. Reichow et al. (Eds.), *Handbook of early childhood special education.* (pp. 3–19). Springer International Publishing.

McWilliam R. (2016). Birth to three early intervention. In Reichow B., Boyd B., Barton E., Odom S. (Eds.), *Handbook of early childhood special education* (pp. 75-88), Springer.

Meho, L. I., & Yang, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar. *Journal of the American society for information science and technology, 58(*13), 2105-2125. https://doi.org/10.1002/asi.20677

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Prisma Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Medicine, 6*(7), e1000097. https://doi.org/10.1016/j.jclinepi.2009.06.006

Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics, 106*(1), 213-228. https://doi.org/10.1007/s11192-015-1765-5

Mosbah-Natanson, S., & Gingras, Y. (2014). The globalization of social sciences? Evidence from a quantitative analysis of 30 years of production, collaboration, and citations in the social sciences (1980–2009). *Current sociology, 62*(5), 626-646. https://doi.org/10.1177/0011392113498866

Oser, C., & Cohen, J. (2003). *Improving Part C Early Intervention: Using what we know about infants and toddlers with disabilities to reauthorize Part C of IDEA.* Retrieved from https://www.zerotothree.org/ztt_policy.html

Otte, E. & Rousseau, R. (2002). Social network analysis: A powerful strategy, also for the information sciences. *Journal of Information Science, 28*(6), 441-453. https://doi.org/10.1177/016555150202800601

Öncül, N. (2015). Review of the Articles Related with Early Childhood Special Education in Turkey. *International Journal of Early Childhood Special Education, 6*(2), 247-284. https://doi.org/10.20489/intjecse.75598

Safford, P., Sargent, M., & Cook, C. (1994). Instructional models in early childhood special education: Origins, issues, and trends. In P. Safford (Ed.), *Early childhood special education: Yearbook in early childhood education* (Vol. 5, pp. 96—117). New York: Teachers College Press.

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science, 24*(4), 265-269. https://doi.org/10.1002/asi.4630240406

Smith, B. J. (2000). The federal role in early childhood special education policy in the next century: The responsibility of the individual. *Topics in Early Childhood Special Education, 20*(1), 7-13. https://doi.org/10.1177/027112140002000102

Snyder, S., & Sheehan, R. (1983). Integrating research in early childhood special education: The use of meta-analysis. *Diagnostique, 9*(1), 12-25. https://doi.org/10.1177/073724778300900102

Van Eck, N. J., & Waltman, L. (2014). Visualizing bibliometric networks. In Y. Ding, R. Rousseau, & D. Wolfram (Eds.), *Measuring Scholarly Impact: Methods and Practice* (pp. 285–320). Springer.

Van Eck, N. J., & Waltman, L. (2017). Citation-based clustering of publications using CitNetExplorer and VOSviewer. *Scientometrics, 111*(2), 1053-1070. https://doi.org/10.1007/s11192-017-2300-7

Virués-Ortega, J. (2010). Applied behavior analytic intervention for autism in early childhood: Meta-analysis, meta-regression and dose–response meta-analysis of multiple outcomes. *Clinical Psychology Review, 30*(4), 387-399. https://doi.org/10.1016/j.cpr.2010.01.008

Wang, J., Peters, H. P., & Guan, J. (2006). Factors influencing knowledge productivity in German research groups: Lessons for developing countries. *Journal of Knowledge Management, 10*(4), 113-126. https://doi.org/10.1108/13673270610679408

Webb, S. J., & Jones, E. J. (2009). Early identification of autism: Early characteristics, onset of symptoms, and diagnostic stability. *Infants and Young Children, 22*(2), 100. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5232420/

White, H. D., & McCain, K. W. (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972–1995. *Journal of the American society for information science, 49*(4), 327-355. https://doi.org/10.1002/(SICI)1097-4571(19980401)49:4%3C327::AID-ASI4%3E3.0.CO;2-4

Wolery, M. (2003). A unified theory of practice in early intervention/early childhood special education: Evidence-based practices. *The Journal of Special Education, 37*(3), 164-173. https://doi.org/10.1177/00224669030370030601

Yildirim, E. S., & Akçamete, G. (2014). Determination of Difficulties Encountered by Mothers Having Children with Multiple Disabilities During Early Childhood Special Education Services Process. *Cumhuriyet Uluslararası Eğitim Dergisi, 3*(1), 74-89. https://doi.org/10.30703/cije.321336

Zitt, M., Bassecoulard, E., & Okubo, Y. (2000). Shadows of the past in international cooperation: Collaboration profiles of the top five producers of science. *Scientometrics, 47*(3), 627-657. https://doi.org/10.1023/a:1005632319799

Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods, 18(*3), 429-472. https://doi.org/10.1177/1094428114562629

Zwaigenbaum, L., Bryson, S., & Garon, N. (2013). Early identification of autism spectrum disorders. *Behavioural Brain Research, 251*, 133-146. https://doi.org/10.1016/j.bbr.2013.04.004

Article Information Form

Evaluating ChatGPT in Generating Feedback on Content and Organization Components of EFL Compare and Contrast Essays

Amine Hatun Ataş[1*] (ID)
Behice Ceyda Cengiz[2] (ID)
Berkan Çelik[3] (ID)

[1] Galatasaray University, Distance Education Application and Research Center, İstanbul, Türkiye, aminehatunatas@gmail.com

[2] Zonguldak Bulent Ecevit University, Faculty of Humanities and Social Sciences, Department of Translation and Interpreting, Zonguldak, Türkiye, behiceceydacengiz@gmail.com

[3] Van Yuzuncu Yil University, Başkale Vocational School, Department of Computer Technologies, Van, Türkiye, berkancx@gmail.com

*Corresponding author

Abstract: ChatGPT, an innovative large language model that has impressed worldwide audiences with its exceptional generative capabilities, is now positioned to significantly transform the field of education. The purpose of this exploratory study is to investigate how accurately ChatGPT generates feedback on the content and organization components of EFL compare and contrast essays and the extent to which the feedback length provided by ChatGPT differs from that of the human teacher.To address these questions, a ChatGPT prompt incorporating evaluation criteria for content and organization components was developed, generating feedback on 10 compare and contrast student essays using the ChatGPT 3.5 version. The ChatGPT feedback and teacher feedback were assessed quantitatively and qualitatively according to the predetermined evaluation criteria. Furthermore, two types of feedback were compared descriptively and by conducting the Wilcoxon Sign Rank Test. The findings revealed that ChatGPT produced highly accurate feedback for both content and organization components, surpassing the teacher in the length of feedback provided. While the accuracy rate of the generated feedback was high, issues such as holistic assessment of the essay, false positives, failure to provide feedback where needed, and discrepancies in the depth of feedback compared to teacher feedback were identified. The results suggest that while ChatGPT shows promise in providing educational feedback, teacher-AI collaboration in giving feedback for EFL compare and contrast essays is important for delivering feedback that optimally benefits learners.

Keywords: Artificial Intelligence, Compare and Contrast Essay, ChatGPT, EFL, Feedback

1. Introduction

Exploring the etymology of the term feedback is advantageous for establishing a precise understanding of the concept. The initial definition of feedback entails the process of redirecting a portion of the output of a machine, system, or process back to its input. This terminology emerged from discussions in the field of electrical engineering and rocket science during the early 20th century (Burke & Pieterick, 2010). Steinmetz (1915) exemplified this concept by illustrating that when a cable is grounded, the current at its end undergoes reversal, flowing back into the cable, thus termed "feeding back" rather than existing from it (as cited in Merriam-Webster, n.d.). Later, in the field of education, Kulhavy (1977) defined feedback simply in his work *Feedback in Written Instruction* as "… any of the numerous procedures that are used to tell a learner if an instructional response is right or wrong" (p. 211). Feedback is a very important component in the field of education, more specifically in English as a Foreign Language (EFL) writing classes with many benefits for second language (L2) learners (Biber et al., 2011; Wilson & Czik, 2016). When L2 learners notice a gap between their current language use (interlanguage) and the target language form upon receiving feedback, they become more aware of their errors. According to the noticing hypothesis, form-focused corrective feedback that enables this noticing facilitates L2 learners' language acquisition processes (Schmidt & Frota, 1986). Other research also shows that L2 feedback that is centered on macrolevel aspects of L2 writing such as content, coherence, and cohesion helps L2 learners improve their writing performance (Bakla, 2020; Elola & Oskoz, 2016). Therefore, it becomes important to provide L2 learners with instructional writing feedback not only in the language component of L2 writings but also in such dimensions as content and organization since writing is a versatile skill and requires dealing with high level concepts such as ideation and style (Zhai & Ma, 2023).

Literature shows that L2 teachers experience some challenges related to providing effective writing feedback (e.g., Dikli & Bleyle, 2014; Fu et al., 2024). One of the biggest challenges is pertinent to the significant amount of time and effort required to offer feedback to students, particularly when dealing with multiple students across various classes (Steiss et al., 2024). Offering individualized feedback and assistance can be demanding and resource-intensive for teachers, particularly when they lack the time or resources to attend to each student's needs effectively (Baskara, 2023; Jackson et al., 2022). This challenge may even discourage some teachers from delivering effective feedback and result in superficial feedback (Noroozi et al., 2023). These challenges motivate L2 teachers and researchers to search for alternative ways of giving writing feedback (Huang, 2023). This need resulted in the development of computer programs designed to offer feedback on writing, known as Automated Writing Evaluation (AWE) systems, such as Grammarly, Pigai, and similar software (Zainurrahman & Rojab, 2024). However, recently, new versions of Generative Artificial Intelligence (GenAI), such as ChatGPT, started to replace and transform AWE systems due to their affordances. Unlike previous models, ChatGPT does not need to be trained on human datasets specific to a task or genre. Additionally, ChatGPT's 3.5 version is currently affordable and readily available for everyone (Steiss et al., 2024).

Recent advancements in AI technology suggest that ChatGPT holds significant potential in writing pedagogy, offering automated feedback on students' L2 writing and supporting teachers in the feedback process. Collaborating with ChatGPT while giving feedback on students' writing can reduce teachers' workload, help prevent the fatigue that comes from correcting numerous student assignments and produce more efficient feedback (Barrot, 2023a; Teng, 2024). However, for effective collaboration between teachers and ChatGPT, it is crucial to first examine key characteristics of ChatGPT's feedback, with accuracy being one of the most important qualities to assess (Steiss et al., 2024). Uncovering ChatGPT's capability to provide accurate feedback can inform teachers of optimal utilization of ChatGPT in their feedback practices.

Research on the use of ChatGPT as a provider of L2 writing feedback has been related to teacher and student perceptions about ChatGPT feedback (e.g., Bok & Cho, 2023; Xiao & Zhi, 2023), and comparison of ChatGPT and teacher feedback in terms of such features as the type of feedback, the level of supportive tone in the feedback, and the clarity of the directions given in the feedback (e.g., Banihashem et al., 2024; Guo & Wang, 2024; Steiss et al., 2024). However, there is a growing need for scrutinizing the accuracy of ChatGPT feedback in EFL writing context, as also highlighted in Guo and Wang's (2024) study. As GenAI chatbots continue to gain prominence, there arises a need for additional research to investigate the quality and accuracy of the generated response by these GenAI chatbots, recognizing the dynamic nature of these evolving technologies (Chaka, 2023). To fill in this research gap, the present study investigates how accurately ChatGPT generates feedback concerning the content and organization components of compare and contrast essays written in the context of a CEFR B1 level EFL Vocabulary and Composition class in an undergraduate Applied English and Translation program in Türkiye. The compare and contrast essay type was selected as the focus of this study due to its emphasis on critical thinking and organizational skills, which are essential components of academic writing at tertiary level. Additionally, this essay type is commonly taught in English for Academic Purposes (EAP) at this level. In order to assess whether ChatGPT, as an educational technology, provides accurate writing feedback evaluations in the classroom, teacher feedback was used as a benchmark for comparison which is also a common method adopted in previous studies (e.g., Mizumoto et al., 2024; Pfau et al., 2023). The rationale behind this comparison is to view ChatGPT as a feedback tool that can complement and enhance teacher feedback instead of serving as a substitute for teacher feedback (Guo & Wang, 2024). Content and organization were chosen as the main feedback focus in the current study - since the accuracy of ChatGPT feedback on linguistic features of L2 texts has already been evidenced in previous studies (e.g., Mizumoto & Eguchi, 2023; Mizumoto et al., 2024; Pfau et al., 2023). It was also noted in earlier studies that even less technologically advanced AWE systems can deal with surface-level errors related to language easily

whereas the capability of these technological tools related to giving feedback on content and organization are still questioned (Guo & Wang, 2024). As the second focus of the study, the length of teacher and ChatGPT feedback is compared. Earlier research on L2 writing demonstrates that feedback amount can be a factor that affects learners' perceptions and uptake of feedback (Thi & Nikolov, 2021; Zhang & Hyland, 2018). The reason for this comparison, therefore, is to provide the baseline for future studies which can analyze the effect of feedback length as a variable that can affect the utilization of ChatGPT feedback by L2 learners (Guo & Wang, 2024).

1.1. Chat generative pre-trained transformer (ChatGPT)

ChatGPT, abbreviated for Chat Generative Pre-Trained Transformer, is an AI-driven conversational agent created by the American startup OpenAI. ChatGPT offers multifunctional capabilities (OpenAI, 2024a). It taps into a vast knowledge base, drawing from diverse sources to generate human-like text, serving as a valuable resource for language input and practice. Unlike traditional databases, ChatGPT's corpus is structured as statistical patterns and associations (Barrot, 2023a). The latest iteration of ChatGPT, derived from the GPT 3.5 model, demonstrates improved proficiency in comprehending natural language, enhanced efficiency and accuracy in addressing inquiries, and increased adaptability (Rudolph et al., 2023; Su et al., 2023).

As an AI-driven technology designed to simulate human intelligence, ChatGPT exhibits exceptional proficiency in a wide range of writing tasks including choosing a topic, establishing the context, creating an outline, drafting the content, and making revisions, often comparable to the capabilities of humans (Barrot, 2023a). ChatGPT provides a lot of affordances in the context of EFL. To exemplify, it can be used for material and assessment generation (Pack & Maloney, 2023), improving writing skills and motivation (Song & Song, 2023), grammar check (Schmidt-Fajlik, 2023), question generation (U. Lee et al., 2023), and feedback (Su et al., 2023). Using GenAI can aid in certain situations, such as when working on early writing drafts or when lacking access to a well-trained educator, given the simplicity of automatic feedback generation with satisfactory quality using ChatGPT (Steiss et al., 2024). Studies approached feedback from different perspectives, including the comparison of the scoring of ChatGPT and human-generated feedback (Steiss et al., 2024), comparison of ChatGPT and teacher-generated feedback (Guo & Wang, 2024), or AI-enabled evaluation (Lee, 2023).

Despite the invaluable affordances, ChatGPT comes with significant limitations. To name a few, language models like ChatGPT have the capability to produce false or incorrect statements, often exhibiting low accuracy in various contexts. ChatGPT might also produce fabricated information instead of producing an "I don't know" response (Meyer et al., 2023). The complexity and depth of the responses can be restricted by the lack of sophisticated or iterative prompt engineering in output generation. In addition, submitting each prompt separately to ChatGPT to prevent the learning from previous prompts may also limit the breadth of its responses even further (Barrett & Pack, 2023). It is known that there is a tendency of AI models to hallucinate. When the essay topics primarily focus on argumentation and critical reflection rather than factual accuracy, these hallucinations are not a concern. Still, this does not change the fact that AI models may struggle with factual correctness in some cases (Herbold et al., 2023). Regarding the limitations with respect to feedback, ChatGPT might use different evaluation criteria than the teachers, which could lead to feedback that does not fit teachers' needs. When ChatGPT does not have adequate background information about the class and students, this could lead to inappropriate feedback (Guo & Wang, 2024). Furthermore, ChatGPT provides longer feedback on average, and this may increase the chances of the error rate related to inaccurate content or knowledge. Occasionally, ChatGPT offers constructive feedback alongside numerous instances of irrelevant or excessive information. ChatGPT might need iterative prompting to refine the outputs although it has great capacity to respond. While not as severe as other challenges previously reported, it can be a struggle for ChatGPT and human evaluators to give feedback for the

high-scoring essays with respect to crucial features prioritization, yet this can be solved via the improved prompting (Steiss et al., 2024).

Before the appearance of large language models (LLMs), educators and researchers have been using several approaches including automated feedback (Barrot, 2023b; Ranalli, 2018), AWE (Link et al., 2022), automated corrective feedback tools (Shadiev & Feng, 2023), or natural language processing (NLP) tools (Wang et al., 2020). LLMs have the capacity to analyze grammar, cohesion, and style all at once, while also offering feedback (Bonner et al., 2023). As an LLM type, ChatGPT is capable of offering versatile feedback automatically across various genres and contexts as it does not necessitate an independent training set like other AWE applications (Steiss et al., 2024). Teachers can leverage these affordances of ChatGPT to generate feedback on student writing (Guo & Wang, 2024). However, effective writing of prompts is needed to harness the full potential of ChatGPT.

1.2. Prompt engineering

Prompt engineering is becoming an essential aspect of understanding generative AI as it progresses to become deeply integrated in every aspect of our lives (Bozkurt, 2024). Guidelines and frameworks were developed to construct effective prompts. Giray (2023) put forward that prompts should encompass distinct tasks, contextual information crucial for task completion, a defined question to address, and specifications outlining the format for generating the response. Particularly, Giray (2023) structured the prompts with the elements "instruction, context, input data, and output indicator" (p. 2630). Spasić and Jankovic (2023) designed their prompts to incorporate the role of the AI model, which defines the persona it adopts while responding, the instruction that guides the model in producing the desired outputs, and seed-words that direct the AI's generated output through specific keywords or phrases. Guo and Wang (2024) suggested including more contextual information, such as language proficiency of students, into the prompts for more personalized feedback generation. OpenAI also provided some guidelines for best practices for prompt engineering. They suggested providing specific, detailed descriptions of the desired context, outcome, length, format, style, and other relevant aspects (OpenAI, 2024b). Lo (2023, p. 1) provided the CLEAR ("Concise, Logical, Explicit, Adaptive, and Reflective") framework as a standard method for creating prompts. Based on that framework, the prompt should be clear and precise as well as structured and coherent. It should include clear output specifications, allow flexibility and customization, and be refined and enhanced via continuous evaluation.

AI can be optimized as a valuable resource for enhancing productivity in delivering quality feedback when specific prompts are formed clearly and precisely (Carlson et al., 2023). Several studies have formed unique prompt structures for their needs in EFL contexts (Bonner et al., 2023; Carlson et al., 2023; Guo & Wang, 2024; Huang, 2023; Pack & Maloney, 2023; Schmidt-Fajlik, 2023; Steiss et al., 2024; Su et al., 2023). We examined these specific prompts that were constructed to provide feedback for students' writing. The prompts constructed to provide feedback on writing in these studies cover various aspects of feedback and assessment for EFL learners' writing. To be more specific, they include instructions for providing specific, actionable feedback on essays, evaluating paragraphs based on given criteria, correcting grammar and mechanics in sentences, and providing suggestions for improvement in writing quality, grammar, spelling, vocabulary, and organization. These prompts emphasize the importance of providing constructive feedback, using examples, and adhering to specific criteria or rubrics for evaluation. To sum up, the most common points of these prompts were the role, context, type of the writing, tone and simple language of the feedback, and evaluation criteria.

1.3. Feedback related studies in EFL essay writing

When scrutinizing studies conducted within the domain of GenAI and feedback in EFL writing, a multitude of findings regarding the benefits and limitations of AI tools emerge. Guo and Wang's (2024)

study focused on assessing how ChatGPT could aid EFL instructors in providing feedback on students' writing, initially by analyzing ChatGPT's ability to generate feedback for EFL students' argumentative essays. The findings demonstrated that ChatGPT generated a substantially greater volume of feedback compared to teachers. Furthermore, whereas teacher feedback primarily concentrated on content and language-related concerns, ChatGPT allocated its attention more evenly across the three feedback areas: content, organization, and language. In a similar vein, Wang et al. (2024) compared ChatGPT and teacher feedback on argumentative essays in terms of feedback accuracy and examined the factors affecting their evaluation. This study demonstrated that ChatGPT and teacher feedback had unique affordances and limitations. It was shown that ChatGPT had a considerable accuracy rate, demonstrating promising capability to give writing feedback. This capability, however, was influenced by the utilization of discourse markers and arguments' length. The limitations of ChatGPT were related to the fact that it limited its feedback to the linguistic form while giving affective feedback. Teacher feedback, on the other hand, was advantageous in terms of teachers' available contextual knowledge about the students' immediate needs and progress supported by their ability to have empathy with their students. Another similar study was conducted by Banihashem et al. (2024) who made a comparison between the quality of the feedback provided by ChatGPT and teachers on argumentative essays. Their study revealed that while ChatGPT feedback was more focused on giving informative feedback about how to write an essay, teacher feedback was centered on locating the problems in the essay. It was also shown that there was not any significant relationship between the essay quality and feedback quality. Another comparison study was carried out by Steiss et al. (2024) who evaluated the quality of ChatGPT and teacher feedback in terms of certain criteria such as being criteria-based, accurate, indicating ways for improvement. In that study, teachers who received comprehensive training in providing writing feedback were found to be more effective than ChatGPT in delivering feedback across all areas, except when it came to feedback based on specific established criteria.

Another body of research looked into the effect of AI-assisted language learning on the development of Chinese EFL learners' writing skills. These studies demonstrated the efficacy of AI-assisted language learning on the improvement of L2 learners' writing (Liu et al., 2021; Song & Song, 2023; Yan, 2023) and writing motivation (Song & Song, 2023). Likewise, Su et al.'s (2023) study on ChatGPT's role in guiding writing suggested that it could assist the learners with developing argumentative writing's structural, dialogical, and linguistic aspects. It was also shown that ChatGPT had competency to provide personalized feedback, evaluate content and organization, as well as analyzing language, and proofreading texts. However, its effectiveness was found to depend on the quality of questions and criteria provided by users.

Regarding the student perceptions about ChatGPT, Bok and Cho (2023) studied college students' views on using ChatGPT for revising paragraphs in an academic writing course. Students found ChatGPT helpful and reliable for feedback, appreciating its instant responses and flexibility. It effectively corrected errors in vocabulary, grammar, and paragraph structure. However, challenges included the lack of error descriptions, unclear feedback, inconsistency in responses, worries about reduced authorship, and doubts about its learning effectiveness. In another study related with student perspectives, Xiao and Zhi (2023) revealed that while Chinese college-level EFL learners viewed ChatGPT as a valuable tool for offering them instant feedback and individualized learning experiences, they were skeptical about the accuracy of the ChatGPT outputs.

The necessity for the current study originates from the increasing focus on how teachers can collaborate with ChatGPT to enhance the L2 writing feedback process (Guo & Wang, 2024). Working alongside ChatGPT has the potential to improve the quality and efficiency of feedback provided to students, supporting teachers in refining their feedback practices and decreasing their heavy workload (Barrot, 2023a; Teng, 2024). However, for collaboration to work well, it is important to evaluate key

aspects of ChatGPT feedback, particularly its accuracy (Steiss et al., 2024) and length of its responses (Thi & Nikolov, 2021; Zhang & Hyland, 2018). Gaining insight into its capabilities will help teachers integrate ChatGPT more effectively into their feedback practices. This study is also significant in the Turkish context where most studies related with the utilization of ChatGPT for providing L2 writing feedback mostly focused on the teacher and student perceptions (e.g., Punar Özçelik & Yangın Ekşi, 2024; Üstünbaş, 2024), paying little attention to the accuracy of ChatGPT feedback. As essay writing is a common component of EAP courses across Türkiye, this study provides results that may influence the practical use of ChatGPT-generated feedback in terms of its accuracy and length for the teachers delivering these courses.

In line with this, the purpose of this study is to evaluate the accuracy of ChatGPT in generating feedback on content and organization components of EFL compare and contrast essays, and to compare the length of feedback provided by ChatGPT with that of human teachers in the same points. Accordingly, the following two research questions (RQs) have been addressed:

RQ1: How accurately does ChatGPT generate feedback on the content and organization of EFL compare and contrast essays?

RQ2: How does the length of feedback given by ChatGPT and teacher on the content and organization components of compare and contrast essays differ and is there a significant difference between the two?

2. Method

2.1. Research design

This study adopted an exploratory research perspective. Effectively exploring a phenomenon requires adopting two key orientations: flexibility, which involves being adaptable in the search for data, and openness, which entails being receptive to various sources for obtaining that data. The emphasis in exploration always lies on inductively generating new concepts and empirical generalizations. During exploration, both quantitative and qualitative data may be collected. While qualitative data often dominate in exploratory studies, they are supplemented with descriptive statistics whenever possible and appropriate (Stebbins, 2001). Our study aims to explore new ground by investigating how accurately ChatGPT generates feedback on the content and organization components of EFL compare and contrast essays and the extent to which the length of the feedback provided by ChatGPT and that of the human teacher differs in EFL context. As this area is not well understood yet, we are using an exploratory research approach to uncover valuable insights. This study will establish a basis for future research examining the direct application of ChatGPT in the classroom, potentially incorporating the student perspective by analyzing the accuracy and length of the feedback generated by ChatGPT in the specified context. Also, this method allows us to be flexible and open in collecting data, helping us to develop new ideas and generalizations through exploration.

2.2. Context and sample

The compare and contrast essays used in the current study were collected from a Vocabulary and Composition course taught by the second author at the department of Foreign Languages and Cultures at a state university in Türkiye. The medium of instruction was English in that department. The writers of the essays enrolled in the afore-mentioned course were 10 Turkish EFL freshmen with CEFR B1 level in English. The essays were randomly selected from this class with homogenous writing proficiency, with all students experiencing difficulty in essay writing due to challenges with grammar and writing skills. According to the short demographic questionnaire that asked about their age, gender, and participation in preparatory school conducted at the beginning of the semester, seven were female and three were male while their ages ranged from 19 to 21. Before their studies at the department, all of these students attended one-year intensive preparatory program where they

practiced the four language skills, including writing, through 20 hours of weekly instruction and reached a B1 level, as evidenced by the proficiency exam conducted at the end of the program. At the preparatory school, they learnt to write different types of texts such as informal e-mails, blog posts and formal letters of complaint. They also received instruction on how to write an opinion essay. The three-hour Vocabulary and Composition course, which lasted 14 weeks, had two hours of instruction allocated for the writing component of the course and included the teaching of different types of essays, such as cause and effect essay, compare and contrast essay, and argumentative essay respectively, in its syllabus. The decision to focus on compare and contrast essays in the current study, rather than on cause and effect essays, which were taught as the first essay type, was made to give learners additional essay-writing practice. This decision was based on the second author's observation, who also teaches the course, that students struggled with essay writing. Before writing the compare and contrast essay, these students received four weeks of instruction and were expected to write their essays during the final week and submit them to the teacher, which accounted for 20% of their final grade. The instruction covered content-related topics, such as writing an effective hook and thesis statement in the introduction paragraph, using examples and explanations as supporting sentences in similarity and difference paragraphs and organization related topics such as unity, coherence, cohesion and the use of appropriate linking words for this essay type. In-class activities included teacher lectures, as well as group and pair work that engaged learners in analyzing sample essays and completing exercises on connectors, punctuation rules, and grammar topics closely related to compare and contrast essays. The essays they were required to write had to consist of four paragraphs, including introduction, similarity, difference and conclusion paragraph. Although this essay type can include four or five paragraphs, the decision to use four was based on the way it was taught in class, due to the difficulty students had with essay writing. To prevent struggles associated with writing essays on unfamiliar subjects and provide them with greater flexibility and comfort, 10 different compare and contrast essay topics were offered in order to allow them to choose one they felt comfortable with.

2.3. Data collection/generation tools

a) Rubric

The rubric used for providing feedback on compare and contrast essays was developed by integrating elements from two main sources. Some of the items were derived from the instructional content related to compare and contrast essays, as outlined by Buitrago and Díaz (2018). Additionally, for some of the items in the organization section of the rubric, an academic writing book titled "Writing to communicate: Paragraphs and Essays" which is written by Boardman and Frydenberg (2002) was consulted. The rubric was made up of sections related with content and organization. Under the content category, there were 3 content-related questions for the introduction paragraph, 4 questions for the similarity paragraph, 4 questions for the difference paragraph and 3 questions for the conclusion paragraph. As for the organization category, there were 5 questions in total which are pertinent to unity, coherence and cohesion aspects. To promote the validity and reliability of the rubric, expert opinion was gained from 2 academicians: one having a PhD degree in English Language Teaching and the other having 15 years of teaching experience in a university context. Both of these academicians also had more than 10 years of experience in teaching and grading college-level essay writing. The rubric was piloted with two essays by these academicians, who later provided some wording suggestions for some items to make them clearer. These suggestions were discussed and incorporated into the final version of the rubric.

The rubric was utilized to assess the accuracy of the feedback provided by ChatGPT (See Appendix A). Researchers added an evaluation range to the rubric criteria for this purpose. A three-point evaluation scale was defined as below:

1: Very poor: The feedback is entirely incorrect/irrelevant.

2: Average: The feedback is correct but not comprehensive enough.

3: Very good: The feedback is entirely accurate and comprehensive.

Additionally, a comment section was added for noting differences between feedback provided by ChatGPT and teacher feedback, facilitating the qualitative analysis.

b) Prompt

This part explains how we structured our prompts to get feedback from ChatGPT. As ChatGPT works based on the given prompts, how a person structures a prompt is strongly linked to the output ChatGPT creates. First, we tried two zero-shot prompts ("Can you provide feedback on the student's essay? /Can you provide feedback on the content and organization of the student's essay?") to observe what feedback ChatGPT generates. While these prompts generated a substantial amount of feedback, it lacked sufficient structure. Then we analyzed the general guidelines suggested for prompt engineering as well as the various prompts formed in EFL studies on feedback to improve the quality of feedback. In light of these studies, two distinct prompts were generated for feedback on content and organization.

Specifically, in order to create the necessary prompt structure, we specified the role (the persona the AI model adopts when responding) (Bonner et al., 2023; Steiss et al., 2024; Su et al., 2023), student level (the proficiency level of the student the AI model is addressing) (Huang, 2023), expectation (the desired outcome or standard the AI model's response should meet) (Huang, 2023; Steiss et al., 2024), tone (the emotional or stylistic approach the AI model uses in its responses) (Steiss et al., 2024), and criteria (the specific standards or metrics for the AI model's output) (Carlson et al., 2023; Pack & Maloney, 2023). We formed few-shot prompts and in the end, specified fine-tuned prompts. For the fine-tuned prompt, we iteratively prompted ChatGPT to obtain the best refined feedback results based on the evaluation criteria we provided. Ultimately, this process resulted in fine-tuned prompts tailored for optimal feedback generation. We refined our last effective prompt to the following:

> As an English language instructor, generate feedback based on the comparison-contrast essay provided in this session. Students' English level is [B1]. Use a friendly and encouraging tone with simple language. If needed, provide examples of how the student could improve the essay. Instead of rewriting the paragraph, give specific examples and guidelines on how to revise. Be clear and specific in your feedback, and try to include as many corrections as possible. While giving feedback, just focus on *[the criteria lists given to you for the introductory, similarity, difference and conclusion paragraphs.] [the organization criteria list given to you.]*

For the complete working prompt, see Appendix B.

2.4. Procedure

To initiate the research process, the essays written previously by students in an essay writing class taught by the second author of the current study were obtained. Later evaluation criteria for these essays were developed according to the instructional content covered in that class. Thus, the essays, evaluation criteria, and prompts served as necessary input required for the feedback to be provided by ChatGPT. For accuracy check, the second author of the study initially provided her own feedback on the 10 essays using the identified evaluation criteria while the other two authors ran ChatGPT sessions for gaining feedback on these essays. The ChatGPT feedback was conducted using April 2024 version of ChatGPT 3.5. Content feedback and organization feedback prompts were run in separate sessions to prevent ChatGPT from learning from sessions and generating unwanted pieces of feedback. After the process of ChatGPT feedback was completed and the feedback was stored in a text document, the accuracy check of those feedback was done by the second author based on the feedback she gave on

the same essays before. That is, the researcher compared her own feedback on each item in the rubric with ChatGPT feedback and gave scores for their accuracy. For reliability purposes, this accuracy analysis was also done by another expert in English Language Teaching (ELT) who used the same rubric to give feedback on the student essays. In relation to the second research question, the length of the ChatGPT feedback was calculated using a word count function.

The time spent generating feedback was also noted to clarify the procedure. Accordingly, the total duration for teacher feedback was 11 hours and 40 minutes, with an average of 1 hour and 10 minutes spent on each essay. Producing feedback for 10 essays via ChatGPT took 2 hours and 13 minutes, averaging 13.3 minutes per essay.

2.5. Data analysis

To answer the first research question, quantitative data obtained from the evaluation rubric was analyzed descriptively. The frequencies of the "Very poor," "Average," and "Very good" categories were reported for each criterion of the content and organization components. Additionally, the mean (M) and standard deviation (SD) values were reported. Furthermore, the feedback produced by ChatGPT was compared to teacher feedback qualitatively. Accordingly, comments generated for each essay were categorized into themes using thematic analysis (Braun & Clarke, 2006) and were further elucidated with relevant quotations.

Regarding the second research question, the word counts of ChatGPT and teacher feedback given for content, organization, and total essays were analyzed descriptively. The minimum (min.), maximum (max.), M, and SD values were reported. To determine if there was a significant difference between the length of teacher feedback and ChatGPT feedback, the non-parametric Wilcoxon Signed Ranks Test was conducted using IBM SPSS Statistics Version 23. In this examination, the dependent variable, feedback length, is at the continuous level, with quantification reliant on word count. The independent variables comprise two matched pairs, which were total ChatGPT feedback length and total teacher feedback length; ChatGPT feedback length on content and teacher feedback on content; ChatGPT feedback length on organization and teacher feedback on organization. Three distinct comparisons were executed employing a single test. Consequently, a Bonferroni correction was implemented to mitigate the effects of multiple comparisons, thereby establishing the significance threshold at .016, which was adjusted from .05 divided by 3.

2.6. Reliability and validity

The criteria list used for generating ChatGPT feedback for the EFL compare and contrast essays, encompassing content and organization aspects, was developed by the second author who is an expert in the field of English language teaching with graduate degree in the ELT department and experience in teaching and researching writing for almost 10 years. The resulting criteria list underwent further scrutiny by two academicians, each with over 10 years of experience in teaching essay writing at college level. These academicians piloted the rubric with two compare and contrast essays to validate its effectiveness. The piloting resulted in revisions in the wording of some items, producing the final version of the rubric.

The ChatGPT prompt was tested and refined by two experts, both with over 10 years of experience in the field of Computer Education and Instructional Technology (CEIT), who are also the authors of this paper. The refinement process was guided by prompt criteria outlined in the literature. The two essays used during the rubric testing were also employed to check the functionality of the prompt. The ChatGPT feedback generated for the two essays was reviewed by the second author, further clarifying the final version of the developed prompt. The ChatGPT feedback was checked as a precaution against the potential risk of generating inaccurate feedback. For reliability of data analysis, accuracy

assessment was conducted by 2 researchers. One of them was the second author of the study while the other assessor was a researcher with a PhD in ELT and taught essay writing at university for over 10 years. To ensure a true understanding of the criteria, the second author provided explanations for each item and guided the other researcher in assessing the accuracy of a separate essay, which was not one of the 10 chosen essays. Then, ten pieces of ChatGPT feedback data were independently evaluated by two EFL experts based on the rubric. The agreement rate of these evaluation sets was computed individually for each essay. Two essays yielded a consensus of 91%, while two others reached 95.5%. The consensus of 100% was attained for the remaining six essays. Instances of discordance were subjected to deliberation by two domain experts in order to reach a consensus. Miles and Huberman (1994) set an 80% agreement level as an acceptable threshold, and the acquired values fulfill this criterion of reliability.

2.7. Ethical Procedure

The ethics committee report for this study was obtained from the Zonguldak Bülent Ecevit University Human Research Ethics Committee with the decision dated 29.05.2014 and numbered 2014/08-13.

3. Findings

3.1. Feedback accuracy

The accuracy of ChatGPT feedback has been evaluated using a three-point scale, taking the teacher feedback, which serves as a reliable benchmark of expert judgment, as a baseline. Regarding the findings related to content feedback, it is found that ChatGPT's feedback aligns nearly with three criteria of the teacher's feedback for the introduction paragraph of the essay. Relatively less alignment of ChatGPT feedback is found in the feedback provided for the conclusion paragraph. Overall, the mean values of the assessments are above 2.50, with a calculated total content mean value of 2.63 (SD = .17) (See Table 1).

Table 1

Feedback Accuracy of ChatGPT Feedback on the Content of the Compare and Contrast Essays

Feedback Accuracy	Introduction Paragraph (IP)			Similarity Paragraph (SP)				Difference Paragraph (DP)				Conclusion Paragraph (CP)		
	IP1	*IP2*	*IP3*	*SP1*	*SP2*	*SP3*	*SP4*	*DP1*	*DP2*	*DP3*	*DP4*	*CP1*	*CP2*	*CP3*
Very poor	0	0	0	2	1	1	2	0	1	1	3	2	1	1
Average	0	0	1	1	1	2	3	3	1	1	0	2	4	3
Very good	10	10	9	7	8	7	5	7	8	8	7	6	5	6
Mean	2.97			2.53				2.63				2.43		
SD	.10			.39				.37				.35		

Note1: A total of 10 essays are evaluated for each criterion.
Note 2: M and SD values are reported for each paragraph.

It has been determined that ChatGPT feedback is 100% accurate regarding the criteria of having an attention-grabbing hook sentence in the essay (IP1) and including background information in the introductory paragraph (IP2) that describes the context of the topics. However, it has been observed that in the criteria of restating the thesis (CP1), summarizing the similarities and differences written in previous paragraphs (CP2), and stating the students' own opinion about the topic in the conclusion

paragraph (CP3), ChatGPT's feedback is less accurate compared to the feedback given for the other paragraphs.

According to the findings related to organization feedback, it has been found that ChatGPT's feedback on the organization of the essay generates mostly accurate feedback in terms of unity, coherence, and cohesion. Particularly, its feedback on the cohesion of the similarity (O3) and difference paragraphs (O4) has been found to be 100% accurate. The lowest accuracy rate was identified in the feedback generated for the unity criterion (O1). In general, the average scores for the accuracy evaluations exceed 2.50, with a computed total mean score for organization of 2.84 (*SD* = .22) (See Table 2).

Table 2

Feedback Accuracy of ChatGPT Feedback on the Organization of the Compare and Contrast Essays

Feedback Accuracy	*O1* *Unity*	*O2* *Coherence*	*O3* *Cohesion of SP*	*O4* *Cohesion of DP*	*O5* *Cohesion of CP*
Very poor	1	0	0	0	0
Average	3	1	0	0	2
Very good	6	9	10	10	8
Mean	2.50	2.90	3.00	3.00	2.80
SD	.70	.31	.00	.00	.42

Note1: A total of 10 essays were evaluated for each criterion.
Note 2: M and SD values are reported for each criterion.

The unity criterion, for which relatively less accurate feedback is generated, pertains to the relevance and consistent adherence to the central theme throughout both individual paragraphs and the entire essay. The coherence of paragraphs discussing similarities (O3) and differences (O4), evaluated with ChatGPT's feedback being entirely accurate, is judged based on the standard of employing appropriate connectors to ensure clear transitions between the sentences.

Overall, according to the qualitative analysis, issues have been identified concerning the feedback provided by ChatGPT on both content and organization, particularly in relation to unity criterion. During the feedback generation process, an inclination has been noted for ChatGPT to concentrate on individual paragraphs, consequently neglecting the central theme of the essay. In essence, it falls short of providing a comprehensive evaluation of the essay. Regarding the depth of feedback, while in some instances, ChatGPT offers more informative feedback compared to teacher feedback, in others, it provides less informative feedback. The issue of non-informativeness was predominantly observed in the feedback provided for similarity and difference paragraphs. Lastly, it is found that ChatGPT is incapable of detecting certain errors. Moreover, there have been observations indicating that ChatGPT erroneously identifies non-existent errors as mistakes, as demonstrated by instances falling under the false positive theme (See Table 3).

Table 3

Qualitative Examination of the Accuracy of Feedback Provided by ChatGPT

Theme	Code	Explanation	Quotation
Unity	NA	It concerns the matter of relevance and the continual preservation of the central theme throughout both individual paragraphs and the entirety of the essay.	**Teacher feedback:** "*The similarities are off-topic and do not express a similarity that job and education shares. Take these similarities out and write new ones, please.*" **ChatGPT feedback**: "*You've mentioned two similarities, taking notes and listening to the teacher, but they could be presented more clearly. For example, "Both classical education and modern learning techniques emphasize the importance of active listening and note-taking in the classroom."* **Feedback rubric criterion:** O3
Depth of feedback	Informativeness	It denotes the extent of feedback depth, offering thorough explanations in instances where a student makes an error.	**Teacher Feedback:** "*The essay does not present ideas in a well-organized way.*" **ChatGPT feedback**: "*Your essay needs better coherence to help your reader follow your ideas more easily. Within each paragraph, your sentences should flow logically from one to the next. For instance, in your "Difference" paragraph, you mention taking notes, then suddenly jump to discussing the importance of the study environment. It would be clearer if you introduced each difference one at a time and provided examples or explanations for each. This way, your reader can understand each point before moving on to the next.*" **Feedback rubric criterion:** O2
	Non-informativeness	It pertains to feedback that is superficially given when a student commits an error	**Teacher Feedback**: "*Use connectors to start the difference paragraph (e.g., Despite similarities), to introduce each difference (The first difference is that, the second difference is that). While describing the differences, please use such connectors as "On the contrary, while, despite, though, whereas". To provide examples, please use such connectors as "For example, For instance "and, "thus, therefore" for explanations.*" **ChatGPT feedback**: "*Your use of connectors to link sentences and paragraphs is limited, which affects the overall clarity of your essay.*" **Feedback rubric criterion**: O4
Capability	No feedback	It denotes the lack of any feedback given when it is necessary.	**Teacher feedback**: "*Your last sentence is not meaningful. Hollywood is not a film. Please, remove it and state your own opinion about the topic.*" ChatGPT could not generate any meaningful feedback. **Feedback criterion**: CP3
	False positive	It denotes a circumstance where feedback mistakenly suggests the existence of an error or characteristic when it is actually not present.	**ChatGPT feedback**: "*There is a need to start the conclusion with a clear transition: "In conclusion, while high school and university share some similarities, they also exhibit significant differences..."* Feedback was not provided as it was deemed unnecessary by the teacher. **Feedback criterion**: O5

Note1: Feedback criteria lists are provided on Appendix A.
Note 2: NA: Not Applicable, O: Organization, CP: Conclusion Paragraph

As presented in Table 3, the qualitative analysis of comments written on the accuracy of ChatGPT feedback revealed three key themes: unity, feedback depth, and capability. Concerning unity theme, as quoted, although a student wrote about similarities in her/his essay, ChatGPT did not correctly identify the relevance of these similarities to the main topic and instead provided feedback suggesting that the presented similarities were just unclear. In terms of feedback depth, which also affects its

length, two main themes emerged: informativeness and lack of informativeness. As seen in the examples presented in the Table 3, while the teacher's feedback notes that the student's ideas were not well-organized, ChatGPT's feedback explains this in greater detail. On the other hand, in some instances where the teacher offered a more detailed explanation, ChatGPT provided only a brief statement in its feedback. Regarding capability theme, we observe that in some instances, although the teacher provided feedback, ChatGPT did not give any feedback on the same points. On the other hand, there are cases where the teacher deemed feedback unnecessary, but ChatGPT still provided feedback.

3.2. Feedback length

The length of feedback provided by both ChatGPT and the teacher was compared based on the word count. The word count of the 10 student essays ranged from a min. of 184 to a max. of 239, with a mean of 216.00 and a SD of 19.26. Table 4 presents the descriptive statistics of feedback length.

Table 4

Feedback Length

Feedback focus	*Min.*	*Max.*	*Mean*	*SD*
ChatGPT feedback length on content	358	718	499.10	93.06
ChatGPT feedback length on organization	304	617	475.60	88.56
Total ChatGPT feedback length	1223	1548	1398.50	115.87
Teacher feedback on content	199	462	343.00	84.31
Teacher feedback on organization	52	246	146.00	56.28
Total teacher feedback length	406	786	611.60	129.51

Note: The feedback length is determined by the number of words.

Wilcoxon Signed Ranks tests were conducted to compare total ChatGPT feedback length and total teacher feedback length, ChatGPT feedback length on content and teacher feedback on content, as well as ChatGPT feedback length on organization and teacher feedback on organization. The tests elicit that ChatGPT created significantly higher length of feedback in three comparisons (Total ChatGPT-Teacher: Z = -2.803, p = .005; Content feedback- ChatGPT-Teacher: Z = -2.599, p = .009; Organization feedback-ChatGPT-Teacher: Z = -2.803, p = .005) at .016 significance level, based on two-tailed tests. Positive ranks were used for the calculation of Z-values (See Table 5).

Table 5

Wilcoxon Signed Ranks Test Statistics

	ChatGPT-Teacher *(Total)*	**ChatGPT-Teacher** *(Content)*	**ChatGPT-Teacher** *(Organization)*
Z	-2.803[a]	-2.599[a]	-2.803[a]
Asymp. Sig. (2-tailed)	.005	.009	.005

Note: a. Based on positive ranks.

4. Discussion

The first research question addressed how accurately ChatGPT generated feedback on the content and organization of EFL compare and contrast essays. The findings of the study demonstrated that the accuracy of ChatGPT feedback on the content and organization components of EFL compare and contrast essays was considerably high for each paragraph. These findings are in line with those of earlier studies. In their study, Banihashem et al. (2024) reported not finding any significant inaccuracy

in the feedback given by ChatGPT on argumentative essays. Likewise, the studies by Wang et al. (2024) and Steiss et al. (2024) substantiated the accuracy of the feedback provided by ChatGPT for argumentative texts. In a similar vein, Su et al. (2023) noted that ChatGPT was considerably competent in giving feedback on the argumentative essays in terms of content and organization aspects.

The feedback provided by ChatGPT on the introduction paragraph was found to be completely accurate in terms of content. However, the content-wise accuracy of feedback in similarity and difference paragraphs was lower than that in the introduction paragraph although those paragraphs still had a high accuracy rate. Additionally, the accuracy of content feedback was relatively lower in the conclusion paragraph than in the other paragraphs. The findings show that when there is a need for the linking of some ideas in different paragraphs, ChatGPT can fail to give accurate feedback since it cannot merge the ideas from different paragraphs effectively and considers the essay as a whole consisting of related parts. This can be considered as a unity problem, which represents ChatGPT's inability to look at the essay holistically. This finding is also evident in the analysis of feedback accuracy in the organization component, which demonstrates that unity is the feedback aspect having the least accuracy rate in that component. This problem was also highlighted in Steiss et al.'s (2024) study which revealed that ChatGPT failed to identify the mistake when a student confused a proper name for another proper name, which signified that ChatGPT did not understand the text as a problem related to unity.

As qualitative data show, ChatGPT occasionally fails to detect if the relevance of the ideas is maintained throughout the essay or not. Other identified distinctions in ChatGPT feedback related to depth of feedback and capability. These findings suggest that while using ChatGPT feedback, teachers should check if unity is achieved through maintenance of the main theme within and across the paragraphs. They also need to examine whether the feedback adequately addresses identified problems in a student essay, and whether additional feedback from the teacher is necessary. Furthermore, whether ChatGPT ignores some mistakes or gives wrong feedback although there is no need for feedback also needs to be checked by teachers. Concerning the capability theme, Guo and Wang's (2024) study also showed that ChatGPT could sometimes give irrelevant feedback. Under the capability theme, the issue related to overlooking necessary feedback was also noted by Wang et al. (2024) who put forward that teacher feedback gave more focused feedback addressing critical problems in essays, which can be ignored by ChatGPT. Related to the informativeness theme, the situations where ChatGPT or teacher gave more in-depth feedback than one another were observed. The latter situation can be explained by ChatGPT's lack of contextual information about the students and their progress as also emphasized in other studies (Guo & Wang, 2024; Wang et al., 2024). The in-depthness of the ChatGPT feedback can be attributed to its tendency to provide more directive feedback, unlike teachers, who, as noted in Guo and Wang's (2024) study, tend to offer more indirect feedback.

As for the second research question which investigated the difference in the length of feedback provided by ChatGPT and teacher, it was shown that ChatGPT provided longer feedback than teachers, which was corroborated by Guo and Wang (2024) and Wang et al. (2024). Considering that ChatGPT has paramount capacity to provide more voluminous and detailed feedback in just a few seconds in comparison to teachers who spend a greater amount of time, it can be stated that ChatGPT proves to be an efficient tool in terms of time and effort required for feedback effort. Therefore, when all the affordances and limitations of ChatGPT are taken into consideration, it can be argued that collaboration between teacher and ChatGPT is required for an optimal integration of ChatGPT feedback in L2 writing classes.

Finally, despite the capacity of LLMs to give feedback by analyzing grammar, cohesion, and style all at once (Bonner et al., 2023), its success is very much dependent on the prompt structure. For taking

advantage of ChatGPT, prompts should be written thoroughly (Carlson et al., 2023). Previous studies have already provided guidance for writing prompts, yet each situation might come with unique requirements and characteristics, and therefore, prompts that are tailored to the specific situations at hand should be ensured. In this way, the above-mentioned issues about the feedback output of ChatGPT can be eased, and this might help educators maximize the effectiveness, accuracy, quality, and practicality of feedback generation by creating their own fine-tuned prompts that align closely with the desired context and criteria.

5. Conclusion

Conclusively, it can be affirmed that ChatGPT delivers exceedingly precise feedback concerning both content and organizational aspects within EFL compare and contrast essays. It is inferred that ChatGPT provides feedback closely resembling teacher feedback for the introduction paragraph in terms of content criteria; nevertheless, it exhibits decreased accuracy in generating feedback for the conclusion paragraph based on the relevant criteria. This could be because assessing the conclusion paragraph requires a comprehensive view of the entire essay, including factors like paraphrasing the thesis statement, summarizing similarities and differences, and ensuring coherence across the paragraphs. ChatGPT's deficiency in maintaining unity in this aspect may have contributed to the relatively lower mean score. When providing feedback based on organizational criteria, it is deduced that ChatGPT performs well in generating feedback on the coherence of similarity and difference paragraphs. However, it produces less accurate feedback when evaluating the essay for unity and providing corresponding feedback. Although the error rate is low, upon examining the errors, it is concluded that ChatGPT excels in paragraph and sentence-level evaluations but encounters difficulties in evaluating the essay holistically. The comparatively greater length of ChatGPT feedback compared to teacher feedback highlights its strength over teacher feedback and indicates a potential solution to the general problem teachers face in providing detailed feedback to all students. The conclusion reached is that there is a necessity for collaboration between teachers and ChatGPT, rather than delivering ChatGPT feedback directly to students, at least with this version of ChatGPT. This suggests that ChatGPT feedback should be reviewed by teachers before being shared with students. This study shows that the precision of ChatGPT feedback is notably elevated contingent upon the prompt criteria, thereby offering guidelines outlining the facets of the generated ChatGPT feedback necessitating scrutiny by educators and researchers.

6. Implications

The results of this study provide a guideline for points to consider before utilizing ChatGPT feedback in research and classroom applications. Prior to providing ChatGPT feedback to students, it should be evaluated whether the feedback comprehensively addresses the essay, provides sufficient explanations, and identifies any areas where feedback is lacking or incorrect. Additionally, the high accuracy rate of ChatGPT feedback obtained in this study indicates that ChatGPT shows promise in providing educational feedback. Diversifying feedback evaluation criteria and conducting in-depth content analysis of generated feedback represent significant areas for research and application. These efforts are crucial for understanding how LLMs interpret written text, identifying areas of difficulty, and determining where they might outperform humans.

This study also provides implications for general prompt engineering and prompt engineering for generating feedback with GenAI tools, especially with ChatGPT. When the text provided to ChatGPT contains any automatic numbering/bullets, ChatGPT fails to recognize these numbers/bullets. Hence, it is advisable to avoid using automatic numbering/bullets and to use manual numbering/bullets for the piece of text intended for ChatGPT. As ChatGPT can pause at times, it is important to compare the generated feedback with the number of criteria provided to ChatGPT. This ensures that ChatGPT's

pauses do not lead to incomplete or insufficient feedback. There is always a risk for ChatGPT to incorporate the exact criteria provided or utilize sample structures from the criteria in its feedback. For this reason, it is crucial to ensure that the generated feedback does not replicate the exact structures from the given criteria. When the prompts are entered separately in different sessions, it generates more detailed feedback. In organization prompts, it is necessary to present the essay as a whole without any paragraph distinctions so that ChatGPT can evaluate the overall organization of the essay effectively. Additionally, it is necessary to connect the sections of the essay that refer to each other. For instance, ChatGPT needs to check the introductory paragraph again to see if the thesis statement is restated in the conclusion section since thesis statements are initially presented in the introductory paragraph of the compare and contrast essay.

7. Limitations and Recommendations

In this study, the accuracy and length of ChatGPT feedback were experimentally analyzed by researchers. Evaluating the effectiveness of ChatGPT feedback from the perspective of students and examining its direct impact on students' essays are areas that remain underexplored and are suggested as future research topics. Additionally, technology-related challenges in utilizing ChatGPT for feedback generation by teachers were not investigated, given the emphasis on research over practical classroom implementation. Future research could involve comparing the teacher-generated feedback with that of ChatGPT feedback in terms of efficiency.

Within this research, the feedback length was calculated by considering the total word count of the content generated by ChatGPT and teachers. In future research, determining feedback idea units and conducting comparisons based on feedback types could be beneficial in understanding the potential of ChatGPT feedback. Concerning LLM model, feedback was generated using the GPT-3.5 version. Comparing the feedback produced by different LLMs could be valuable in understanding the potential of models in providing educational feedback.

This study produced feedback specifically tailored for compare and contrast type essays. It is suggested that further research examines the accuracy and effectiveness of AI-generated feedback for various other essay types within the EFL context. Further studies can also focus on the prompt structures and provide a comparative analysis of generated feedback by different prompt structures. Finally, this study sampled 10 student essays. Working with a larger sample could more clearly elucidate potential issues/strengths in ChatGPT's feedback generation.

References

Bakla, A. (2020). A mixed-methods study of feedback modes in EFL writing. *Language Learning & Technology, 24*(1), 107–128. https://doi.org/10125/44712

Banihashem, S, K., Kerman, N.T., Noroozi, O., Moon, J., & Drachsler, H. (2024). Feedback sources in essay writing: Peergenerated or AI-generated feedback? I*nternational Journal of Educational Technology in Higher Education, 21*(23), 1-15. https://doi.org/10.1186/s41239-024-00455-4

Barrett, A., & Pack, A. (2023). Not quite eye to AI: Student and teacher perspectives on the use of generative artificial intelligence in the writing process. *International Journal of Educational Technology in Higher Education*, *20*(59). https://doi.org/10.1186/s41239-023-00427-0

Barrot, J. S. (2023a). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing, 57.* https://doi.org/10.1016/j.asw.2023.100745

Barrot, J. S. (2023b). Using automated written corrective feedback in the writing classrooms:

Effects on L2 writing accuracy. *Computer Assisted Language Learning*, *36*(4), 584-607. https://doi.org/10.1080/09588221.2021.1936071

Baskara, F. (2023). Integrating ChatGPT into EFL writing instruction: Benefits and challenges. *International Journal of Education and Learning, 5*(1), 44-55. https://doi.org/10.31763/ijele.v5i1.858

Biber, D., Nekrasova, T., & Horn, B. (2011). The effectiveness of feedback for L1-English and L2-writing development: A meta-analysis. *ETS Research Report Series, 2011*(1), i-99. https://doi.org/10.1002/j.2333-8504.2011.tb02241.x

Boardman, C.A., & Frydenberg, J. (2002). *Writing to communicate: Paragraphs and Essay* (2nd ed.). Pearson Education.

Bok, E., & Cho, Y. (2023). Examining Korean EFL College Students' Experiences and Perceptions of Using ChatGPT as a Writing Revision Tool. *Journal of English Teaching through Movies and Media, 24*(4), 15-27. https://doi.org/10.16875/stem.2023.24.4.15

Bonner, E., Lege, R., & Frazier, E. (2023). Large language model-based artificial intelligence in the language classroom: Practical ideas for teaching. *Teaching English with Technology*, *23*(1), 23-41. https://doi.org/10.56297/BKAM1691/WIEO1749

Bozkurt, A. (2024). Tell me your prompts and I will make them true: The alchemy of prompt engineering and generative AI. *Open Praxis, 16*(2), 111-118. https://doi.org/10.55982/openpraxis.16.2.661

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Buitrago, C. R., & Diaz, J. (2018). Flipping your writing lessons: Optimizing time in your EFL writing classroom. In J. Mehring & A. Leis (Eds.), *Innovations in flipping the language classroom.* (pp. 69–91) Springer.

Burke, D. M., & Pieterick, J. (2010). *Giving Students Effective Written Feedback*. Maidenhead: Open University Press.

Carlson, M., Pack, A., & Escalante, J. (2023). Utilizing OpenAI's GPT-4 for written feedback. *TESOL Journal*, *759*, e759. https://doi.org/10.1002/tesj.759

Chaka, C. (2023). Generative AI chatbots-ChatGPT versus YouChat versus Chatsonic: Use cases of

selected areas of applied English language studies. *International Journal of Learning, Teaching and Educational Research*, *22*(6), 1-19. https://doi.org/10.26803/ijlter.22.6.1

Dikli, S., & Bleyle, S. (2014). Automated essay scoring feedback for second language writers: How does it compare to instructor feedback? *Assessing Writing, 22*, 1–17. https://doi.org/10.1016/j.asw.2014.03.006

Elola, I., & Oskoz, A. (2016). Supporting second language writing using multimodal feedback. *Foreign Language Annals, 49*(1), 58-74. https://doi.org/10.1111/flan.12183

Fu, Q. K., Zou, D., Xie, H., & Cheng, G. (2024). A review of AWE feedback: types, learning outcomes, and implications. *Computer Assisted Language Learning, 37*(1–2), 179–221. https://doi.org/10.1080/09588221.2022.2033787

Giray, L. (2023). Prompt engineering with ChatGPT: A guide for academic writers. *Annals of Biomedical Engineering*, *51*, 2629-2633. https://doi.org/10.1007/s10439-023-03272-4

Guo, K., & Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies, 29,* 8435-8463*.* https://doi.org/10.1007/s10639-023-12146-0

Herbold, S., Hautli-Janisz, A., Heuer, U., Kikteva, Z., & Trautsch, A. (2023). A large-scale comparison of human-written versus ChatGPT-generated essays. *Scientific Reports*, *13*(1), 18617. https://doi.org/10.1038/s41598-023-45644-9

Huang, J. (2023). Engineering ChatGPT Prompts for EFL Writing Classes. *International Journal of TESOL Studies, 5*(4), 73-79. https://doi.org/10.58304/ijts.20230405

Jackson, D., Davidson, P.M., & Usher, K. (2022). *Feeding Back and Feeding Forward*. In: Successful Doctoral Training in Nursing and Health Sciences. Springer, Cham. https://doi.org/10.1007/978-3-030-87946-4_5

Köroğlu, & A. Çakır (Ed.), *Fostering foreign language teaching and learning environments with contemporary technologies*. (pp. 115-133). IGI Global. https://doi.org/10.4018/979-8-3693-0353-5.ch006

Kulhavy, R. W. (1977). Feedback in written instruction. *Review of Educational Research, 47*(2), 211–232. https://doi.org/10.3102/00346543047002211

Lee, A. V. Y. (2023). Supporting students' generation of feedback in large-scale online course with artificial intelligence-enabled evaluation. *Studies in Educational Evaluation*, *77*, 101250. https://doi.org/10.1016/j.stueduc.2023.101250

Lee, U., Jung, H., Jeon, Y., Sohn, Y., Hwang, W., Moon, J., & Kim, H. (2023). Few-shot is enough: exploring ChatGPT prompt engineering method for automatic question generation in English education. *Education and Information Technologies*, 1-33. https://doi.org/10.1007/s10639-023-12249-8

Link, S., Mehrzad, M., & Rahimi, M. (2022). Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Computer Assisted Language Learning*, *35*(4), 605-634. https://doi.org/10.1080/09588221.2020.1743323

Lo, L. S. (2023). The CLEAR path: A framework for enhancing information literacy through prompt engineering. *The Journal of Academic Librarianship*, *49*(4), 102720. https://doi.org/10.1016/j.acalib.2023.102720

Liu, C., Hou, J., Tu, Y. F., Wang, Y., & Hwang, G. J. (2021). Incorporating a reflective thinking promoting mechanism into artificial intelligence-supported English writing environments. *Interactive Learning Environments, 31,* 3340–3359. https://doi.org/10.1080/10494820.2021.2012812

Merriam-Webster. (n.d.). Get Looped in on 'Feedback' In Merriam-Webster.com dictionary. Retrieved from https://www.merriam-webster.com/wordplay/the-history-of-feedback#:~:text=Feedback%2C%20which%20began%20as%20an,of%20coming%20out%20of%20it

Meyer, J. G., Urbanowicz, R. J., Martin, P. C., O'Connor, K., Li, R., Peng, P. C., ..., & Moore, J. H. (2023). ChatGPT and large language models in academia: opportunities and challenges. *BioData Mining*, *16*(1), 20. https://doi.org/10.1186/s13040-023-00339-9

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Sage.

Mizumoto, A., Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics, 2*(2), 100050. https://doi.org/10.1016/j.rmal.2023.100050

Mizumoto, A., Shintani, N., Sasaki, M., & Teng, M. F. (2024). Testing the viability of ChatGPT as a companion in L2 writing accuracy assessment. *Research Methods in Applied Linguistics, 3*(2), 100116. https://doi.org/10.1016/j.rmal.2024.100116

Noroozi, O., Banihashem, S. K., Taghizadeh Kerman, N., Parvaneh Akhteh Khaneh, M., Babayi, M., Ashrafi, H., & Biemans, H. J. (2023). Gender differences in students' argumentative essay writing, peer review performance and uptake in online learning environments. *Interactive Learning Environments, 31* (10), 6302–6316. https://doi.org/10.1080/10494820.2022.2034887

OpenAI. (2024a). What is ChatGPT? Retrieved from https://help.openai.com/ en/articles/ 6783457-what- is- chatgpt

OpenAI. (2024b). Best practices for prompt engineering with the OpenAI API. Retrieved from https://help.openai.com/en/articles/6654000-best-practices-for-prompt-engineering-with-the- openai-api

Pack, A., & Maloney, J. (2023). Potential affordances of generative AI in language education: Demonstrations and an evaluative framework. *Teaching English with Technology, 23*(2), 4-24. https://doi.org/10.56297/BUKA4060/VRRO1747

Pfau, A., Polio, C., & Xu, Y. (2023). Exploring the potential of ChatGPT in assessing L2 writing accuracy for research purposes. *Research methods in Applied Linguistics, 2*(3), 100083. https://doi.org/10.1016/j.rmal.2023.100083

Punar Özçelik, N., & Yangın Ekşi, G. (2024). Cultivating writing skills: the role of ChatGPT as a learning assistant – a case study. *Smart Learning Environments, 11*(10), 1-18. https://doi.org/10.1186/s40561-024-00296-8

Ranalli, J. (2018). Automated written corrective feedback: How well can students make use of it?. *Computer Assisted Language Learning*, *31*(7), 653-674. https://doi.org/10.1080/09588221.2018.1428994

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching, 6*(1), https://doi.org/10.37074/jalt.2023.6.1.9

Schmidt-Fajlik, R. (2023). Chatgpt as a grammar checker for Japanese English language learners: A comparison with grammarly and prowritingaid. *AsiaCALL Online Journal*, *14*(1), 105-119. https://doi.org/10.54855/acoj.231417

Schmidt, R., & Frota, S. N. (1986). Developing basic conversational ability in a second language: A case study of an adult learner of Portuguese. In R. R. Day (Ed.), *Talking to learn: Conversation in second language acquisition* (pp. 237-326). Newbury House

Shadiev, R., & Feng, Y. (2023). Using automated corrective feedback tools in language learning: A review study. *Interactive Learning Environments*, 1-29. https://doi.org/10.1080/10494820.2022.2153145

Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, *14*, 1260843. https://doi.org/10.3389/fpsyg.2023.1260843

Spasić, A. J., & Janković, D. S. (2023, June 29-July 1). *Using ChatGPT standard prompt engineering techniques in lesson preparation: Role instructions and seed-word prompts* [Paper presentation]. 58th International Scientific Conference on Information Communication and Energy Systems and Technologies (ICEST), Nis, Serbia. https://doi.org/10.1109/ICEST58410.2023.10187269

Stebbins, R. A. (2001). *Exploratory research in the social sciences* (Vol. 48). Sage.

Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., Moon, Y., Tseng, W., Warschauer, M., & Olson, C. B. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction, 91*, 101894. https://doi.org/10.1016/j.learninstruc.2024.101894

Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing, 57,* 100752. https://doi.org/10.1016/j.asw.2023.100752

Teng. M. F. (2024). A systematic review of ChatGPT for English as a foreign language writing: Opportunities, challenges and recommendations. *International Journal of TESOL studies, 6*(3), 36-57. https://doi.org/10.58304/ijts.20240304

Thi, N. K., & Nikolov, M. (2021). Feedback Treatments, Writing Tasks, and Accuracy Measures: A Critical Review of Research on Written Corrective Feedback. *Tesl-Ej*, *25*(3), n3. Retrieved from https://files.eric.ed.gov/fulltext/EJ1332267.pdf

Üstünbaş, Ü. (2024). EFL Learners' Views About the Use of Artificial Intelligence in Giving Corrective Feedback on Writing: A Case Study. In *Fostering Foreign Language Teaching and Learning Environments With Contemporary Technologies* (pp. 115-133). IGI Global.

Wang, D., Su, J., & Yu, H. (2020). Feature extraction and analysis of natural language processing for deep learning English language. *IEEE Access*, *8*, 46335-46345. https://doi.org/10.1109/ACCESS.2020.2974101

Wang, L., Chen, X., & Wang, C., Xu, L., Shadiev, R., & Li, Y. (2024). ChatGPT's capabilities in providing feedback on undergraduate students' argumentation: A case study. *Thinking Skills and Creativity, 51*, 101440. 46335-46345. https://doi.org/10.1016/j.tsc.2023.101440

Wilson, J., & Czik, A. (2016). Automated essay evaluation software in English Language Arts classrooms: Effects on teacher feedback, student motivation, and writing quality. *Computers & Education*, *100*, 94–109. https://doi.org/10.1016/j.compedu.2016.05.004

Xiao, Y., & Zhi, Y. (2023). An Exploratory Study of EFL Learners' Use of ChatGPT for Language Learning Tasks: Experience and Perceptions. *Languages, 8*(3), 212. https://doi.org/10.3390/languages8030212

Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies, 28*, 13943–13967. https://doi.org/10.1007/s10639-023-11742-4

Zainurrahman, & Rojab, S. R. (2024). Examining Bing AI as a Solution to EFL Writing Feedback Challenges. *PROJECT (Professional Journal of English Education), 7*(2). Retrieved from https://journal.ikipsiliwangi.ac.id/index.php/project/article/view/21639

Zhai, N., & Ma, X. (2023). The Effectiveness of Automated Writing Evaluation on Writing Quality: A Meta-Analysis. *Journal of Educational Computing Research, 61*(4), 875-900. https://doi.org/10.1177/07356331221127300

Zhang, Z., & Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing, 36*, 90-102. https://doi.org/10.1016/j.asw.2018.02.004

Article Information Form

Appendix

Appendix A. Rubric for Assessing Content and Organizational Structure in Compare and Contrast Essays

Compare and Contrast Essay Content Feedback Criteria List				
Feedback Criteria *(Introduction paragraph-IP)*	**Assess the accuracy the feedback provided on a scale of 1-3.**			**Comment**
	1	**2**	**3**	
IP1-Is there an engaging hook in the introduction paragraph that grabs the reader's interest?				
IP2-Does the introduction paragraph include background information that contextualizes the topic being discussed?				
IP3-Does the introduction paragraph utilize an expression to present the thesis statement? *(e.g., This essay is written in order to … The purpose of this essay is to… This essay aims at ____(ing)… This essay compares and contrasts… This essay discusses…)*				
Feedback Criteria *(Similarity paragraph-SP)*	**1**	**2**	**3**	**Comment**
SP1-Does the similarity paragraph begin with a topic sentence?				
SP2- Does the similarity paragraph present two similarities about the selected topics and compare them?				
SP3-Does the similarity paragraph incorporate examples and/or explanations to uphold the two similarities?				
SP4-Does the similarity paragraph conclude with a summarizing sentence including the two similarities?				
Feedback Criteria *(Difference paragraph-DP)*	**1**	**2**	**3**	**Comment**
DP1-Is there a topic sentence in the difference paragraph?				
DP2-Does the difference paragraph present two differences about the selected topics and contrast them?				
DP3-Does the difference paragraph incorporate examples and/or explanations to uphold the two differences?				
DP4-Does the difference paragraph conclude with a summarizing sentence including the two differences?				
Feedback Criteria *(Conclusion paragraph-CP)*	**1**	**2**	**3**	**Comment**

	1	2	3	Comment
CP1-Does the conclusion paragraph provide a restatement of the thesis statement?				
CP2-Does the conclusion paragraph provide a summary of the similarities and differences discussed in the similarity and difference paragraphs?				
CP3-Does the conclusion paragraph express the student's personal viewpoint on the topic?				

Compare and Contrast Essay Organization Feedback Criteria List				
Feedback Criteria (O: Organization) (O1: Unity; O2: Coherence; O3, O4, O5: Cohesion)	**1**	**2**	**3**	**Comment**
O1-How well does the essay maintain unity? *Unity: This pertains to the issue of relevance and the consistent maintenance of the central theme within both individual paragraphs and the entirety of the essay. Unity within a paragraph is achieved when the supporting sentences enhance comprehension of the main point introduced at the paragraph's outset.*				
O2-How well does the essay maintain coherence? *Coherence: This pertains to the logical progression and linking of ideas within a sentence, the connection between sentences (the transitions between them) within a paragraph, and the continuity across paragraphs.*				
O3: Does the similarity paragraph incorporate connectors to begin the similarity paragraph, present each similarity, offer examples/explanations and provide a concluding statement? *(To begin with, the first similarity is..., for example, for instance, the second similarity is that..., also, as well as, as, both, most important, likewise/like, in the same manner /way, same/similar/similarly, the same as, too, in brief)*				
O4: Does the difference paragraph incorporate connectors to begin the difference paragraph, present each difference, contrast each topic, offer examples/explanations and provide a concluding statement? *(The first difference is that..., on the contrary, while, despite, though, whereas, for example, for instance, thus, therefore, in brief)*				
O5: Does the conclusion paragraph incorporate connectors to begin the paragraph, mention similarities, differences and state the student's personal opinion on the topic? *(In conclusion, as a result, to conclude, to sum up, both, in addition, on the contrary, as far as I am concerned, to my view, it is my impression that, from my point of view)*				

Appendix B. Prompts

Content prompt

As an English language instructor, generate feedback based on the comparison-contrast essay provided in this session. Students' English level is B1. Use a friendly and encouraging tone with simple language. If needed, provide examples of how the student could improve the essay. Instead of rewriting the paragraph, give specific examples and guidelines on how to revise. Be clear and specific in your feedback, and try to include as many corrections as possible. While giving feedback, just focus on the criteria lists given to you for the introductory, similarity, difference and conclusion paragraphs.

Introductory paragraph: "*paste here*"
Similarity paragraph: "*paste here*"
Difference paragraph: "*paste here*"
Conclusion paragraph: "*paste here*"
Introductory paragraph criteria list: "*paste here*"
Similarity paragraph criteria list: "*paste here*"
Difference paragraph criteria list: "*paste here*"
Conclusion paragraph criteria list: "*paste here*"

Organization prompt

As an English language instructor, generate feedback based on the comparison-contrast essay provided in this session. Students' English level is B1. Use a friendly and encouraging tone with simple language. If needed, provide examples of how the student could improve the essay. Instead of rewriting the paragraph, give specific examples and guidelines on how to revise. Be clear and specific in your feedback, and try to include as many corrections as possible. While giving feedback, just focus on the organization criteria list given to you.

Organization criteria list: "*paste here*"
Essay: "*paste here*"

Note: The criteria list can be found in Appendix A. The complete research data and prompts are available upon request from the corresponding author.

Examination of Research Conducted on the Use of Artificial Intelligence in Science Education

Faruk Arıcı [ID]

Bayburt University, Bayburt Faculty of Education, Department of Basic Education, Bayburt, Türkiye, farukarici@bayburt.edu.tr

Abstract: The advancement of artificial intelligence (AI) has been significantly driven by developments in machine learning and neural networks. As AI becomes increasingly pervasive, its applications are diversifying, with notable penetration in sectors such as health, education, social media, robotics, and entertainment. One area in which AI is being deployed is science education. The objective of this study is to examine the research that incorporates AI within the field of science education. By analyzing trends in the reviewed studies, this research identifies the countries, institutions, journals and scholars that are the most prominent contributors to this field of enquiry. The findings suggest that the incorporation of artificial intelligence into science education is still in its infancy, with a paucity of widespread implementation. However, there is a discernible increase in the quantity of published works, with an emerging emphasis on the assessment of learning outcomes and the enhancement of academic performance. The findings indicate that the United States is the leading country in terms of publications related to AI in science education, accounting for 38% of the total contributions. Additionally, Türkiye has emerged as a notable contributor in this field, demonstrating a growing presence. The Journal of Science Education and Technology was identified as the preeminent journal publishing research on AI. Furthermore, the findings revealed that GPT was the most frequently utilized tool in this context. In light of these findings, it is recommended that future investigations into the application of artificial intelligence (AI) in science education employ a range of AI tools and explore the development of higher-order thinking skills.

Keywords: Artificial Intelligence, Bibliometric Analysis Science Education, Science Learning

1. Introduction

The development of artificial intelligence (AI), accompanied by the advent of technologies such as neural networks and machine learning (Wang, 2019), represents a significant advance in computing. In contrast, conventional AI is typically designed to address a specific set of challenges by reacting to distinct inputs and relying on existing datasets (Cheung et al., 2024; Page et al., 2018; Schlegel & Uenal, 2021), Conversely, generative artificial intelligence (GenAI) is engineered to produce novel content or data by recognizing patterns gleaned from training data (Dwivedi et al., 2023). While traditional AI is typically focused on analysis and forecasting, GenAI has the potential to generate novel content, thereby opening up new avenues for creativity and innovation (Abbott & Rothman, 2022). This shift indicates a transition towards an educational framework and societal landscape where AI technology increasingly aids individuals in addressing their challenges (Cooper et al., 2024). GenAI represents a valuable tool for fostering creativity, personalized learning, and inventive pedagogical approaches (Cheung et al., 2024; Perera & Lankathilake, 2023; García-Peñalvo et al., 2023). Through autonomous content creation, GenAI can furnish tailored learning resources, formulate practical exercises, and devise realistic scenarios or simulations for hands-on learning experiences (Mello et al., 2023). The advancement of this technology signals the advent of a new era of innovation, which will redefine the concept of intelligence by embracing the dynamic integration of human ingenuity and machine capabilities (Cooper & Tang, 2024).

In addition to the potential advantages offered by these technologies, there are also significant disadvantages and ethical concerns that must be considered (Mumtaz et al., 2024). Learning Analytics

platforms have the capacity to provide detailed insights into student engagement, performance trends and learning processes. Nevertheless, the Council of Europe (2022) has identified a number of critical ethical issues pertaining to the potential impact of these technologies on the representation of students and teachers. In this context, existing research indicates that interactions with automated artificial systems may result in a reduction in the sense of control experienced by human operators (Berberian et al., 2012; Yusuf et al., 2024). Such a dependency may have a negative impact on human interaction, as it could lead to a decline in the quality of decision-making processes. Additionally, the gathering of personal and confidential data, such as mood analysis, gives rise to substantial ethical concerns, including the monitoring of activities pertaining to political views, ethnic identity, and sexual orientation (Tundrea, 2020). It is therefore of paramount importance to safeguard the privacy of individuals and guarantee the security of data during the utilization of such technologies. The effective utilization of educational technologies will be contingent upon a comprehensive consideration of their potential benefits, in addition to a rigorous addressing of the aforementioned ethical concerns. Furthermore, there is a substantial concern that the advent of new technologies may exacerbate existing inequalities in education, potentially exposing learners to biases and cultural barriers (Malik et al., 2024; Miao et al., 2021). This situation has the potential to negatively impact both the learning process and assessment strategies. For example, while pedagogical agents are designed to enhance educational experiences by emulating human interactions and cultivating positive emotions, they also give rise to concerns pertaining to emotional privacy and the character of virtual relationships (Hudlicka, 2016; Vučković & Sikimić, 2024). Consequently, ethical issues may emerge from competing interests that extend beyond individual users to their families. Discussions surrounding these matters frequently reflect the diverse viewpoints of various stakeholders, whose perspectives differ based on their associations with private companies, government bodies, research institutions, universities, or educational organizations (Popenici & Kerr, 2017). Kowalczuk-Waledziak et al. (2019) posit that the extant framework for teacher education is in need of reassessment, particularly at the nexus of ethical concerns and pedagogical practices. A participatory approach that actively involves educators could facilitate significant advancements in pedagogy (Chichekian & Benteux, 2022). Furthermore, a systematic review of the literature addressing the ethical dimensions of AI applications has revealed a notable lack of focus on ethical considerations, even in the context of the launch of various educational initiatives in schools (Mouta et al., 2023). Notwithstanding the existing challenges, a number of recent initiatives have been launched with the objective of addressing these pressing concerns and needs (Malik et al., 2024). Among these is the UNESCO AI Competence Framework for Educators, scheduled for introduction in 2024. The framework comprises a three-phase progression plan, which is designed to foster a human-centered approach, address ethical considerations related to AI, facilitate comprehension of the fundamental principles and applications of AI, integrate it into pedagogical practices, and utilize it for ongoing professional development (UNESCO, 2023). To illustrate, the AI4T initiative, which seeks to improve the training of educators and school administrators in the field of artificial intelligence, has been developed in a collaborative manner by France, Slovenia, Italy, Ireland, and Luxembourg. The ethical implications of AI in educational contexts have been identified as a key priority within the shared learning objectives established for professional development pathways across these nations. In this context, it is imperative that ethical challenges are effectively addressed in order to create a more inclusive and equitable educational landscape (Mouta et al., 2024).

A review of the last decade reveals that artificial intelligence (AI) has become increasingly pervasive and has exerted a significant influence across a range of domains, including health, education, social media, robotics, and entertainment (Su & Yang, 2022) AI technologies such as GPT-4 and Google Brain have been demonstrated to possess capabilities approaching those of humans, encompassing visual recognition, the ability to write programming code, and the capacity to solve mathematical problems (Cheung et al., 2024). Consequently, as AI has gained a significant presence in society, researchers have

emphasized the importance of improving students' AI literacy (Druga et al., 2019; Ng et al., 2022). Some scientists define AI literacy as comprising cognitive skills and emotional attitudes (Ng et al., 2021). These skills include the ability to comprehend, utilize, and assess AI, as well as to address ethical concerns. While these components align with cognitive skill levels in Bloom's taxonomy (Bloom et al., 1956), they also relate to values associated with AI technologies. However, these mental functions and emotional feelings regarding AI technologies conflict with epistemic considerations and call into question the accuracy of the information produced by the technology. Indeed, the majority of studies concentrate on knowledge and skills (Li & Ironsi, 2024). It would be beneficial to investigate which other variables are examined in the use of AI in the field of SE in order to provide guidance for future studies in this area and highlight potential gaps in the existing literature (Almasri, 2024). It is therefore important to determine the purposes of using AI in SE, the most examined variables, the authors and journals with high impact value, and other trends on this issue in the studies to be conducted. Furthermore, it is essential to investigate the potential effects of using AI in the field of SE in order to gain insight into the current status of AI in SE. Furthermore, the existing literature on the integration of AI in SE can provide insights into the effectiveness of this integration, which can in turn inform curriculum planning and teaching processes in science education (Riera-Negre et al., 2024). Consequently, this research can serve as a pioneering investigation into the present state of AI in SE, offering insights to inform the work of researchers, practitioners, and policymakers by identifying emerging trends. A bibliometric analysis on AI in SE is of particular importance given the rapid development of both fields. A comprehensive and systematic review of the extant literature is a crucial undertaking, as the growing deployment of AI in SE has the potential to usher in transformative shifts in the landscape of science pedagogy (Akhmadieva et al., 2023). The present study employs bibliometric methods to identify areas of prospective development and innovation, ascertain pivotal research trends, and furnish objective, data-driven insights into the current state of integration of AI in SE. In conclusion, given the rapidly increasing volume of publications in the field of SE, bibliometric analysis studies on AI in SE can be stated as an indispensable method to organize this accumulation, to obtain meaningful results, and to identify the most important studies that can guide educational practices and policy-making processes. Furthermore, as Heeg and Avraamidou (2023) have noted, such studies in the literature provide guidance for readers and researchers alike, illuminating potential avenues for future inquiry and underscoring the value of continued investigation in this domain. To illustrate the significance of this endeavor, it is essential to reference the review studies that delineate the evolving trends in research on AI in the field of SE.

When reviewing the existing literature on AI implementation in SE, Cheung et al. (2024) focused on investigating studies utilizing AI, particularly their treatment of epistemological insight. They noted a general need for more reference to this concept within the studies surveyed. Meanwhile, Jia et al. (2023) conducted a study delving into AI within SE, explicitly concentrating on experimental studies from 2013 to 2023. Their inquiry primarily centered on primary and secondary education levels, scrutinizing only those studies with full texts available. Their findings revealed a prevailing superficial integration of AI into SE, particularly noticeable at these educational tiers. The superficial integration of artificial intelligence (AI) has also permeated the domain of educational assessment, where its full potential remains largely untapped. A significant deficiency in the existing literature is the urgent necessity for the development of a comprehensive curriculum framework that effectively integrates AI into the processes of science education assessment. By contrast, Tahiru (2021) examined a wide range of academic subjects, encompassing disciplines such as language education (Liang et al., 2021) and catering to diverse target audiences, including students in higher education settings (Ouyang et al., 2022). It is particularly noteworthy that only one systematic review has explicitly focused on the application of AI within the context of STEM education (Xu & Ouyang, 2022). This paucity of comprehensive investigation suggests a substantial avenue for further inquiry in this pivotal domain. Xu and Ouyang (2022) have identified several noteworthy trends in relation to teaching methodologies,

contextual factors, and the level of educator involvement required to integrate AI within STEM education. The research conducted by Heeg and Avraamidou (2023) sought to answer three key questions: which AI application was utilized, which method was favored, and what was the impact of AI on the outcome. To this end, the researchers analyzed 22 studies. Akhmadieva et al. (2023) conducted a bibliometric analysis similar to the one presented here, but unlike our study, they limited their examination to the Scopus database. Moreover, their analysis of studies conducted between 2002 and 2023 diverged from our study in terms of both temporal scope and the specific studies included. The dissimilarity of the research questions posed by this study and our own study serves to differentiate the two studies from one another. This study will contribute to the existing body of knowledge in this field by employing bibliometric analysis. This analytical approach differs from previous studies in terms of the timeframe, sample selection criteria, methodological approach, and research objectives. The application of bibliometric analysis enables researchers to gain a comprehensive overview of their current scholarly position and to identify new avenues for future research. Furthermore, employing clustering techniques facilitates the identification of emerging publications, notable authors, and prominent journal groups, thereby offering valuable insights into the current landscape of AI research within SE. The research aimed to reveal trends in AI in SE by examining studies on AI in SE using bibliometric analysis. By examining the systematic literature review studies in the literature, it can be seen that the research in question differs from the above mentioned studies in terms of article inclusion criteria, time period of the studies examined, sample size, analysis technique, and specific research questions. Therefore, it can be said that the study will contribute to the literature in accordance with these characteristics. In conclusion, the present study is distinct from existing literature in this field, offering novel insights and contributions to the existing body of knowledge. This research aims to determine the research trends regarding the use of AI in SE by revealing the current situation in the literature and guiding researchers in their studies. The research questions of the study carried out for these purposes are as follows:

1. What is the yearly trend in AI publications within this area?

2. What is the ranking of the leading institutions and countries in research in this field?

3. What are the most frequently addressed topics concerning sustainable development goals?

4. What are the most preferred keywords in this field?

5. What are the terms used extensively in article summaries on AI?

6. Who are the leading authors in this field in the use of AI based on citation and co-citation analysis?

7. Which journals are the most frequently cited for articles on AI in SE, as determined by citation and co-citation analyses?

2. Method

The research was conducted to examine the trends of the studies conducted in Web of Science (WoS) for the use of AI in science education by systematic literature review. The research examined 80 articles on AI in SE in WoS with bibliometric analysis. The bibliometric technique involves applying quantitative techniques on research data such as citations publications and, authors (Broadus, 1987, Pritchard, 1969). Clustering techniques are effective in identifying trending relevant journals, publications and authors thus guiding researchers about the current status of the literature (Van Eck & Waltman, 2017). Bibliometric analysis is essential for determining trends in the literature on a particular research topic (Donthu et al., 2021; Falagas et al., 2006; Moral-Muñoz, et al., 2020; Özbey & Arıcı, 2024; Song et al., 2019). Bibliometric analysis is a technique that reveals the output of a specific research topic in a certain

period with quantitative data (Chen et al., 2019). For this reason, this method was used to obtain numerical data on research trends in examining studies on AI in SE. The search included studies from the date of the first publication of the article on AI in SE to the date of this research (Last access date: 23 April 2024).

2.1. Article selection process

The articles to be examined in the study were analyzed using diagrams that explained the inclusion and exclusion of articles, as is common practice in review studies. The studies were identified using the PRISMA flow diagram (Moher et al., 2012), which is used as a publication for systematic reviews and meta-analyses. Additionally, the diagram utilized by Arici et al. (2019) was employed to identify the studies to be examined. Figure 1, which illustrates the review process and the stages of determining the included and excluded studies, is presented below.

Figure 1

Stages of Determining the Articles Included in the Research
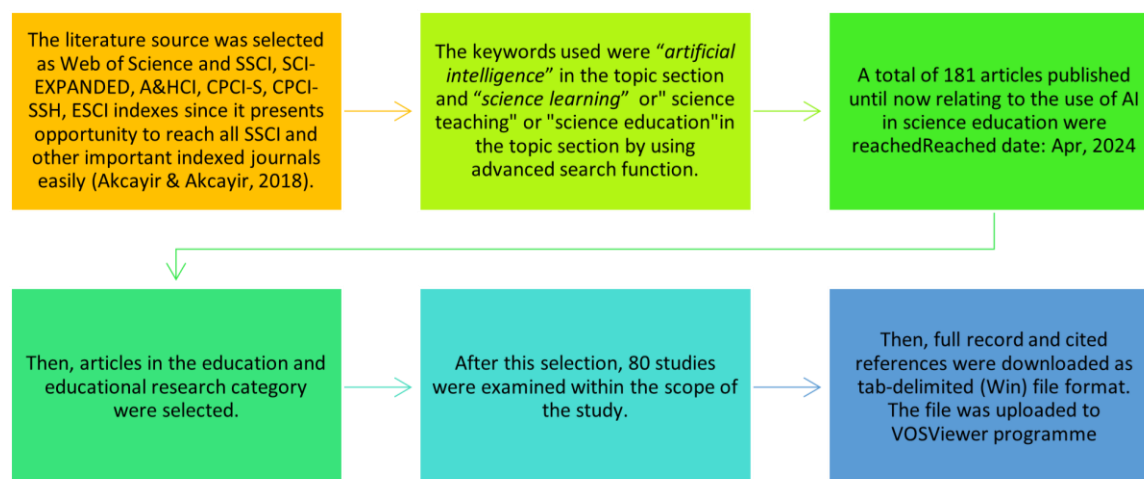
2.2. Validity and reliability

2.2.1. The reliability of the data set

The data set utilized in the analysis was procured from a reliable and widely accepted academic database, such as WoS or Scopus. As such databases furnish data from reputable, peer-reviewed sources, the reliability of our dataset is considerable (Eker et al., 2019; Smith et al., 2015).

2.2.2. Validity of the data collection process

The keywords and search strategies were meticulously devised, and terms that are commonly utilized in literature reviews on the subject were selected. This process ensured the selection of articles that were pertinent to the research question of the study (Mukherjee et al., 2022).

2.2.3. Consistency of analysis methods

The methods most commonly employed in bibliometric analysis, including citation analysis and collaboration network analysis, were applied in accordance with established procedures. These methods represent established analytical techniques that have been validated in previous studies and are widely accepted in the literature (Van Eck & Waltman, 2010).

2.2.4. The reliability of the Results

To guarantee the reliability of the results, the findings were cross-checked with different analytical methods. For example, citation network analysis and word frequency analysis served to validate the findings in different dimensions (Zupic & Čater, 2015).

The aforementioned measures ensure that the study adheres to the requisite standards of validity and reliability. Furthermore, to enhance the representativeness of the data, a comprehensive time span was selected, spanning from the date of publication of the inaugural article to the date of the study's implementation (April 23, 2024). This comprehensive approach enhances the contribution of our study to the existing literature.

3. Findings

3.1. Publication countries/regions

Research data obtained from WoS and showing which countries the researchers examined belong to revealed that among these countries, the USA is the country with the most publications, 38%. Table 1 shows the publication regions of other studies examined within the scope of the study, which is presented below.

Table 1

Frequencies of Articles by Country

Ranking	*Countries*	*Frequencies of Articles*
1	USA	31
2	Spain	7
3	Germany	6
4	China	5
5	Türkiye	4
6	Australia	3
7	Brazil	3
8	Italy	3
9	Sweden	3
10	Canada	2

The results obtained in the analysis carried out to determine the widespread use of AI-related research in the field of SE among countries are presented in Figure 2.

Figure 2

Examination of Articles on AI by Country in the Field of SE


3.2. Publication years

When WOS data is examined, it is seen that the first study on the use of AI in the field of AI was published in 1985. It has been understood that the number of publications is increasing, and the year in which the most publications were made is 2023. The chart showing the publication range of the studies is presented in Figure 3.

Figure 3

Density of Studies According to the Years They Were Published
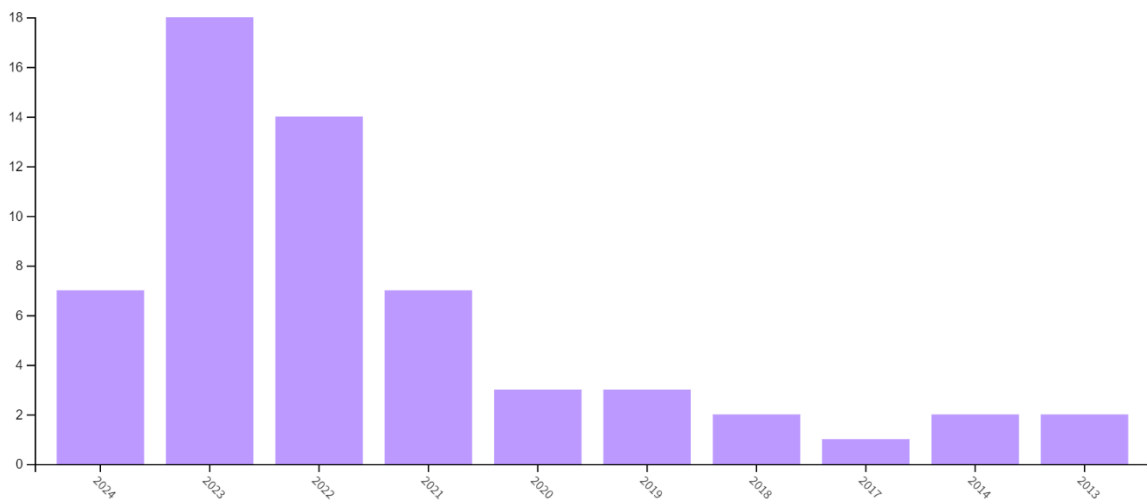
3.3. Institutions where research is conducted

The results obtained from the analyzes carried out to determine the prevalence of AI-related research conducted in the field of AI among institutions, institutes and universities are as shown in Figure 4.

Figure 4

Institutions that Publish the Most Articles on AI in the Field of SE

As illustrated in Figure 4, the University System of Georgia (F=6), the University of Georgia (F=5), Stanford University (F=4), and Curtin University (F=3) are the four institutions with the highest scores.

3.4. Sustainable development goals

Figure 5 illustrates the distribution of research on utilizing AI in SE across Sustainable development goals.

Figure 5

*The Sustainable Development Goals are Discussed in Articles on AI in SE.*

According to Figure 5, it is seen that Quality Education (F=47), Good Health and well-being (F=3), Responsible Consumption and Production (F=2), and No Poverty (f=1).

3.5. Frequently preferred keywords in articles on the use of AI technology in SE

The terms author keywords and co-occurrence were chosen to determine the most recurring keywords in studies on the use of AI in the field of SE. Number 3 was chosen as the frequency of repetition of the keyword, and it was seen that ten keywords matched it. Figure 5 below shows keywords and the relationships between them.

Figure 6

Common Keywords in AI-related Articles in SE
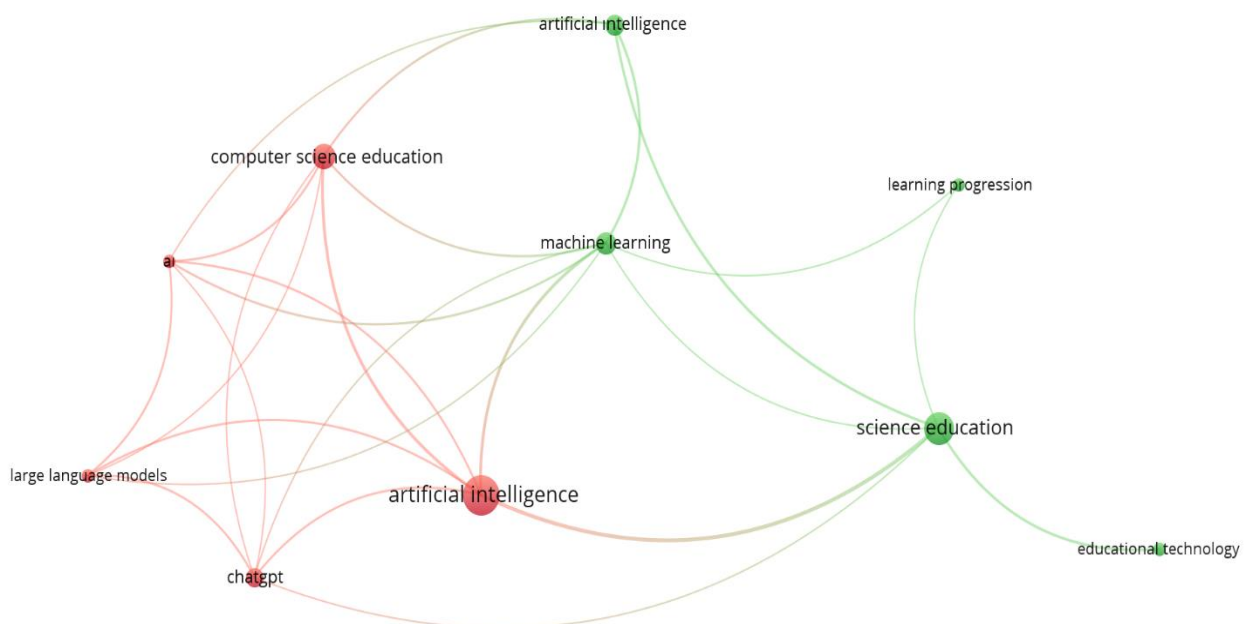
Based on Figure 6, the analysis delineated two distinct clusters, wherein the term "artificial intelligence" emerged as the most prevalent (f=27). Additionally, the examination underscores that keywords such as "SE" (f=18), "computer science education" (f=11), "machine learning" (f=9), "Chatgpt" (f=7), and "educational technology" (f=3) are among the most frequently employed terms in articles pertaining to AI within SE. These findings suggest a predominant focus on science education, computer science education, and machine learning within the articles. A selection of the most commonly occurring keywords identified in the analysis is shown in Table 2.

Table 2

Frequencies of Commonly Occurring Keywords in AI Related Articles in SE

Ranking	Keyword	f	Overall Connection Capacity
1	Artificial Intelligence	27	20
2	Science Education	18	15
3	Computer Science Education	11	12
4	Machine Learning	9	15
5	ChatGPT	7	8
6	Educational Technology	3	3
7	Large Language Model	3	8
8	Learning Progression	3	2

In the analysis carried out to examine the change in the frequency of preference of keywords over time in studies using AI in the area of SE, it was seen that ChatGpt and the Large Language Model have been intensively studied recently. Figure 7 illustrates the progression in the number of articles published on these topics over time. This distribution provides insight into the growing scientific interest in these areas, particularly in the context of AI in SE.

Figure 7

Status of the Keywords Used in the Articles Over the Years
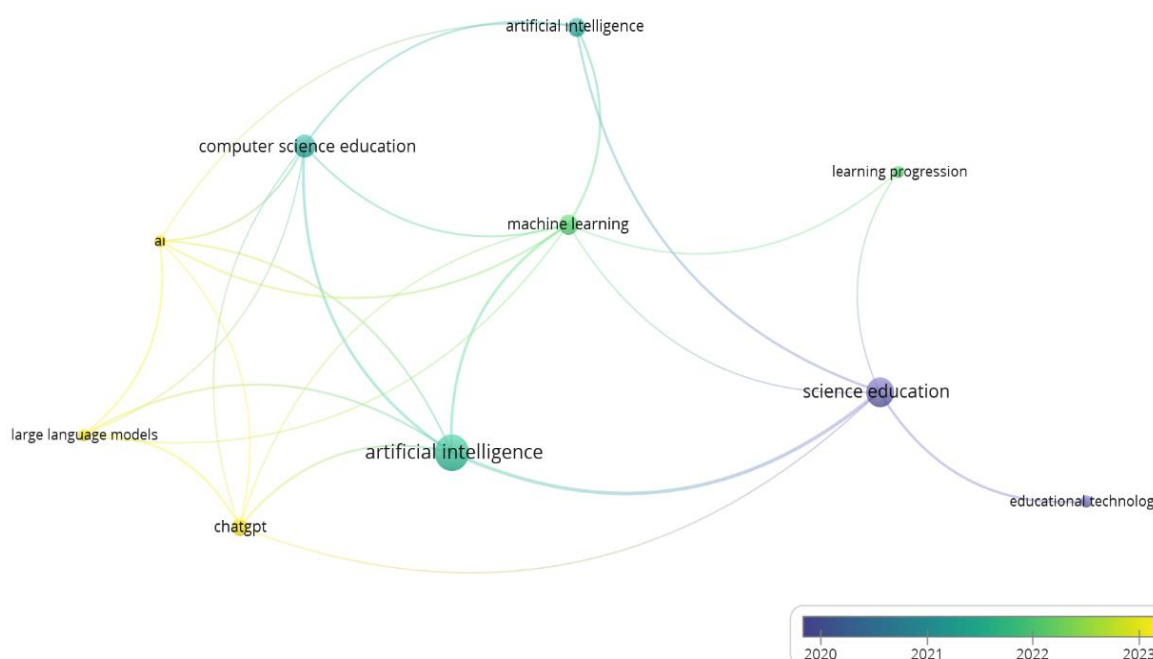
3.6. Repetitive words in research abstracts

In order to determine the frequently repeated words in the abstracts of the studies and thus the variables subject to the studies, the data set was run in the analysis programme. The programme investigated the abstracts of the studies by means of the binary counting method and generated the results. The minimum occurrence threshold for each term was set at 10 and 20 terms were automatically selected for inclusion. Figure 8 below shows the distribution of this data.

Figure 8

Trends in Frequently Used Words in Research Abstracts
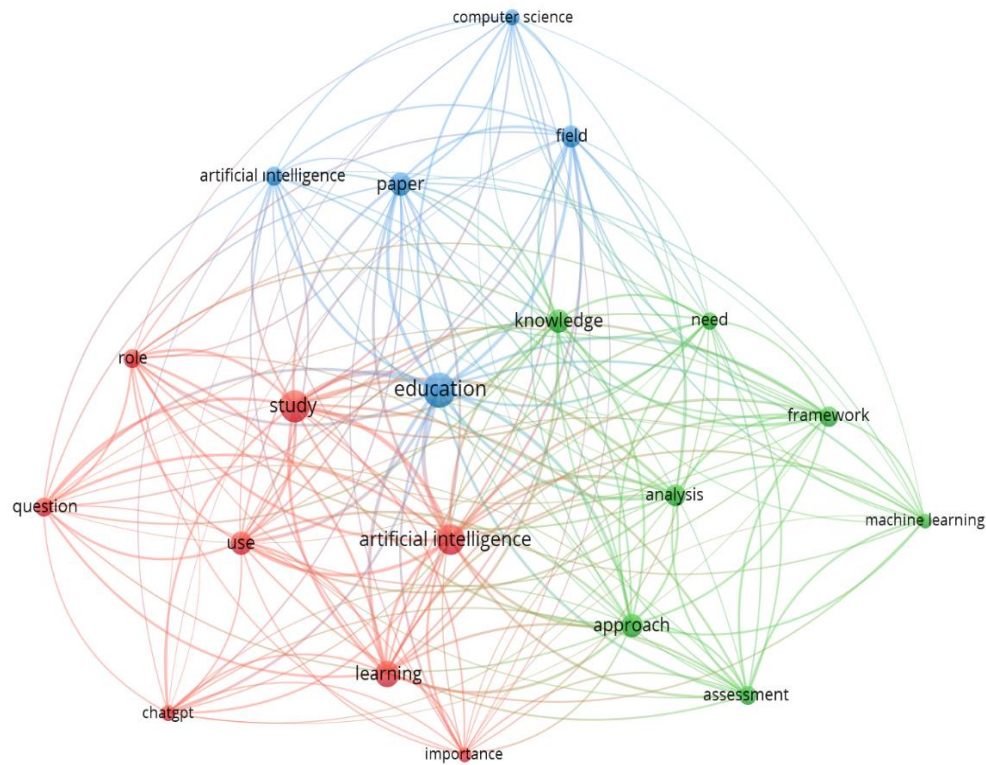
Three clusters of twenty items are shown in Figure 8. In addition, Table 3 provides overviews of commonly repeated words in the abstracts of the articles.

Table 3

*Frequencies of Certain Words in the Study Summaries*

Ranking	Words in Abstract	f
1	Education	52
2	Study	41
3	Artificial Intelligence	40
4	Learning	29
5	Paper	24
6	Knowledge	23
7	Use	23
8	Approach	23
9	Analysis	21
10	Field	20

In Table 3, "education" is the most frequently utilized word in abstracts (f=52), suggesting that it is the dominant concept in articles related to AI in SE. Additionally, the data indicates that "study," "artificial intelligence," "learning," "paper," and "knowledge" are among the most prevalent keywords, with frequencies of f=41, f=40, f=29, f=24, and f=23, respectively (see Table 3).

The distribution of these words over the years shows that the articles focus mainly on evaluation, approach and, more recently, ChatGPT. Figure 9 shows the distribution by year of the most frequently used words in the abstracts of the articles.

Figure 9

Distribution of Most Used Words in Abstracts of Researches by Years
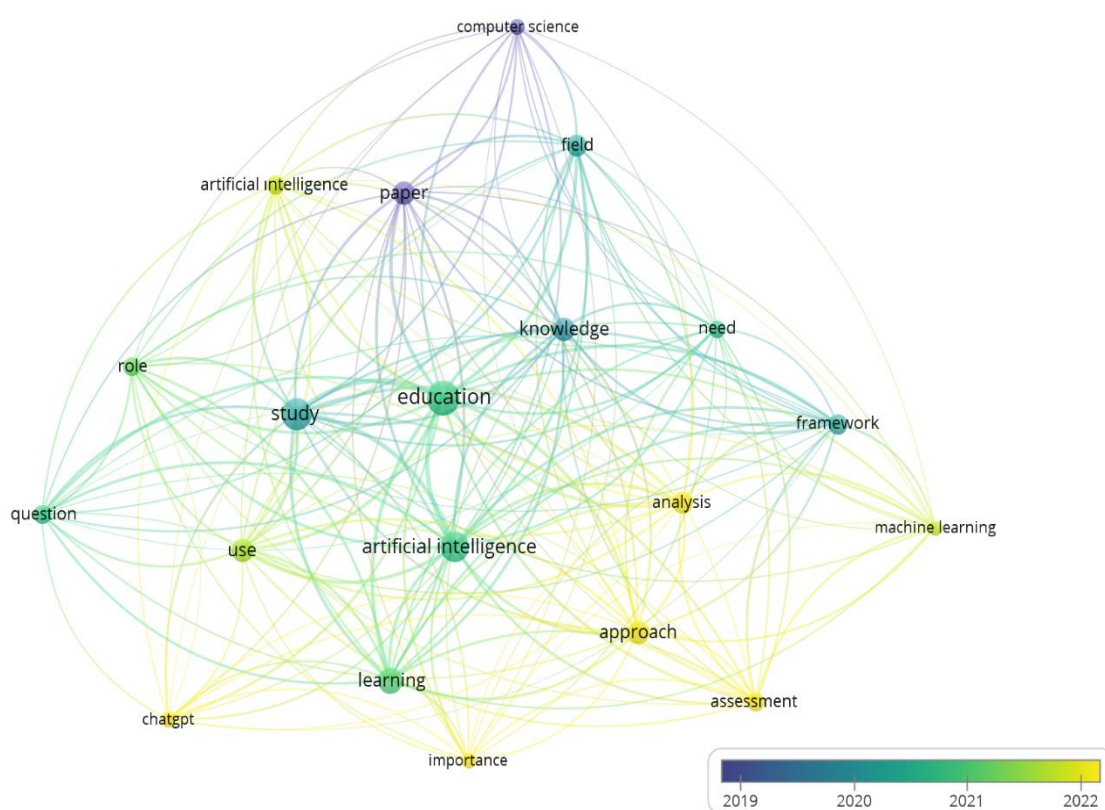
3.7. The most cited authors

Data were analyzed to identify the most cited authors in studies on the use of AI in the field of SE. In order to be included in the review, it is preferred that an author have at least two articles and at least two citations. 8 writers who met this condition were selected by the program. The resulting graph is shown in Figure 10 below.

Figure 10

Authors With the Largest Number of Citations

Figure 10 depicts two clusters and five links. Moreover, a summary of the authors included in the citation analysis is provided in Table 4 below.
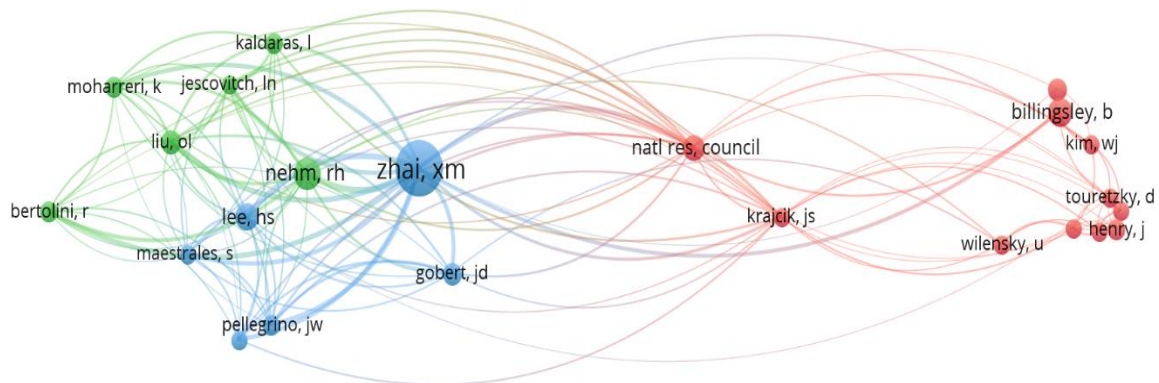
Table 4

Summarised Results of the Most Frequently Cited Authors

Ranking	Authors	Documents	Citations
1	Cooper, Grant	3	106
2	Chin, Doris B	2	72
3	Dohmen, Isla M	2	72
4	Schwartz, Daniel L	2	72
5	Zhai, Xiaoming	4	71
6	Nehm, Ross H	3	32
7	Haudek, Kevin C	2	5
8	Kaldaras, Leonora	2	5

Table 4 shows the number of citations for authors within the field, with Cooper Grant (106 citations), Chin Doris B (72citations), Dohmen, Isla M (72citations) and Schwartz, Daniel L (72citations) as the most cited authors.

A co-citation analysis was conducted utilizing the cited author method, with the minimum number of citations attributable to an author set at six. Twenty-three authors were identified as eligible for inclusion based on the specified criteria. This selection process is illustrated in Figure 11, which depicts the resulting graph of the analysis.

Figure 11

Common Citation Frequency of Top Authors

In Figure 11 there are five clusters and forty items with a total of 592 links. Table 6 below also provides summaries of the authors for the co-citation study.

Table 5

Summary of Top Citations (Co-Citation Analysis)

Ranking	*Authors*	Citations
1	Zhai X M	45
2	Nehm, R H	15
3	Blingsley, B	11
4	Lee, H S	12
5	National Research Council	10
6	Liu, OL	9
7	Gobert, JD	8
8	Cooper, G	8
9	Pellegrino, JW	7
10	Beltorini, R	7

As shown in Table 5, Zhai (45 citations), Nehm (15 citations), Billingsley (11 citations) and Lee (f=12) are the most frequently citing authors in the area.

3.8. Frequently cited journals

To create a visualization illustrating the highly referenced journals, a citation analysis was performed, emphasizing the origins of the publications. The minimum requirement for the count of documents associated with a source was set at 2, while the minimum threshold for citation count was established at 2. The system automatically determined the inclusion of 9 sources. The resultant visualization is depicted in Figure 12.
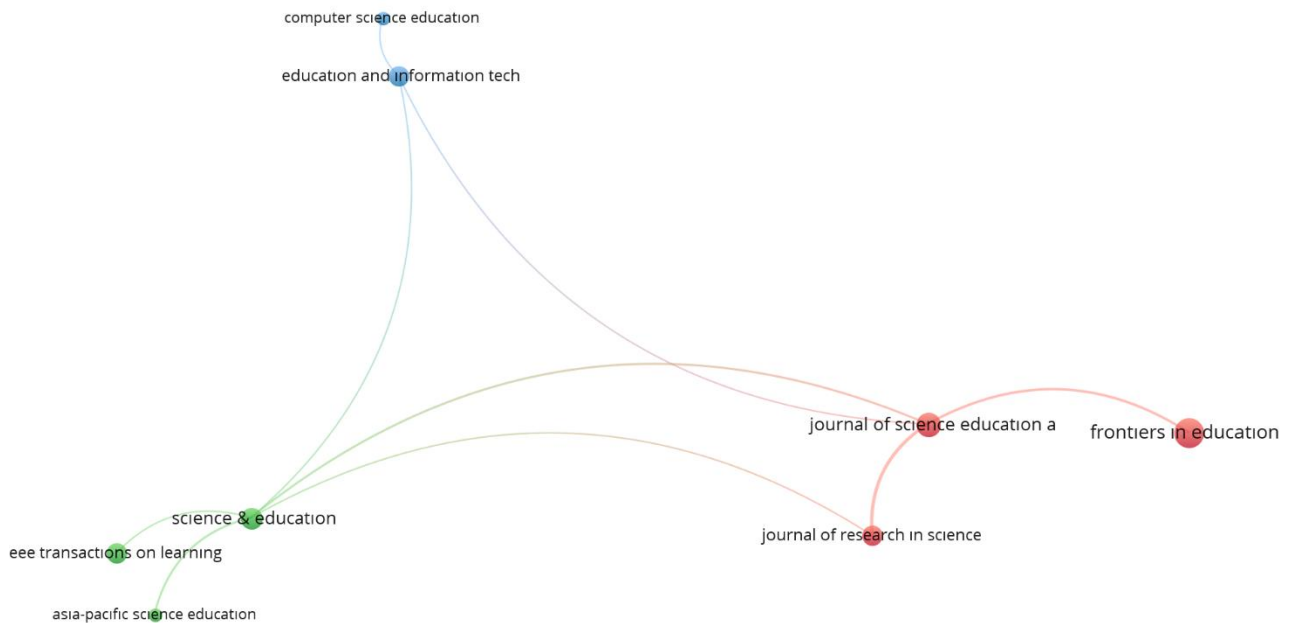
Figure 12

Citation Analysis of the Journals

Figure 12 illustrates the presence of three clusters and nine entries across seventeen links. Moreover, Table 6 below presents a summary of the most frequently cited journals, as determined by a citation analysis.

Table 6

A Summarization of the Most Frequently Cited Journals

Ranking	*Journals*	Documents	Citations
1	Journal of Science Education and Technology	6	155
2	Educational Technology & Society	2	80
3	IEEE Transactions on Learning Technologies	6	54
4	Education and Information Technologies	4	51
5	Journal of Research in Science Teaching	4	33
6	Computer Science Education	2	20
7	Frontiers in Education	9	12
8	Science & Education	5	9
9	Asia-Pacific Science Education	2	5

Table 6 illustrates the journals with the greatest number of citations. These include the Journal of Science Education and Technology (155 citations, 6 documents), Educational Technology & Society (80 citations, 2 documents), IEEE Transactions on Learning Technologies (54 citations, 6 documents), and Education and Information Technologies (51 citations, 4 documents).

In order to conduct the aforementioned co-citation analysis, a visual representation of the sources most commonly referenced in conjunction with one another was produced using the cited sources method. This involved the identification of the sources that were most frequently cited in conjunction with

others. A minimum citation threshold of 15 was established, resulting in the automatic selection of 24 sources for inclusion. The resulting visualization is depicted in Figure 13.

Figure 13

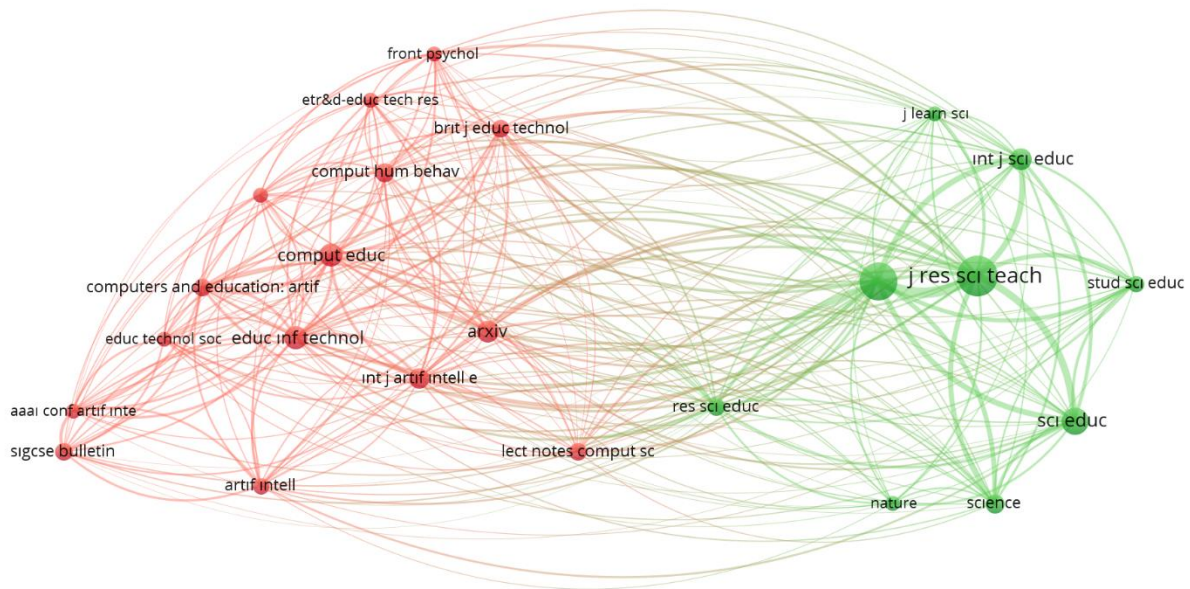*Summary of Highly Referenced Journals*

Figure 13 illustrates the presence of two clusters and a total of twenty-two items at six thousand one hundred fifty-two links. Furthermore, Table 8 presents a summary of the journals that have been cited most frequently (co-citation analysis).

Table 7

Summary of Highly Referenced Journals

Rating Rank	*Journals*	Citations
1	Journal of Research in Science Teaching	108
2	Journal of Science Education and Technology	97
3	Science & Education	50
4	Education and Information Technologies	36
5	Computers & Education	35
6	Arxiv	34
7	International Journal of Science Education	33
8	International Journal of Artificial Intelligence in Education	30
9	Science	25
10	Computers in Human Behavior	25

As demonstrated by Table 7, the four most frequently cited journals in the field are Journal of Research in Science Teaching (f=108), Journal of Science Education and Technology (f=97), Science & Education (f=50), and Education and Information Technologies (f=36).

4. Conclusion, Discussion and Recommendations

To ascertain the utilization of AI in SE, a bibliometric research methodology was employed. This entailed the implementation of a bibliometric mapping analysis, the findings of which revealed valuable insights. The research yielded several important insights regarding the use of AI in the field of social engineering. The results of the research indicate that approximately 38% of the relevant studies were conducted in the USA. The United States of America was followed by Spain, Germany, China, and Türkiye. The frequency of AI-related publications in these countries can be attributed to their endorsement of policies that facilitate the integration of technology in educational settings. The implementation of various policies to popularize the use of technology in these countries provides an explanation for this situation (Eşiyok & Demircioğlu, 2022). Similarly, Akhmadieva et al. (2023) analyzed studies in the Scopus database from 2002 to 2023 in the field of science education and revealed a ranking of the top three countries as the USA, the United Kingdom, and China. It was observed that Spain and the United Kingdom were assigned disparate rankings in the two studies. This discrepancy can be attributed to the utilization of disparate databases and varying timeframes. In the study conducted by Almasri (2024), the United States was ranked second, while Germany and Türkiye were ranked third and fourth, respectively. The following ranking is based on the time period between 2014 and 2023. The present study encompasses the entire time period under consideration. The results demonstrate that Türkiye and Germany have recently allocated greater resources to the development of AI, whereas the United States has maintained a consistent trajectory of advancement in this field. The results of the research indicate that the number of AI-related articles published annually has been increasing continuously since 1985, when the inaugural study was published. The year 2023 saw the greatest number of publications. The slope of the graph shows that this increase will continue. It can be predicted that the use of AI in SE will increase, especially with the emergence of new AI tools and their widespread use (Oh and Lee, 2024). Heeg and Avraamidou (2023) observed that the frequency of publications increased in conjunction with the advent of diverse AI applications that facilitated the advancement of research findings. Another result obtained from the research showed that the leading institutions that distribute research on the use of AI in SE on an institute and school basis are the University System of Georgia, University of Georgia, Stanford University, and Curtin University. It can be said that these universities are trending towards this field by foreseeing the future contributions of AI. As a matter of fact, since the limits of AI cannot be predicted, its contributions cannot be estimated in terms of reaching a conclusion about the extent to which it will be reached. It is an undeniable situation that especially higher education institutions tend to act quickly in terms of focusing on this technology in order to be among the leading institutions in this field (Cheung et al., 2024). In their 2023 study, Ng and colleagues elucidated this phenomenon by highlighting the more advantageous resources and infrastructure available at higher education institutions. The authors asserted that the absence of AI applications in alignment with those of other educational institutions, coupled with their limited access to AI or certain of its facilities, has positioned higher education institutions as the vanguard in this domain among all levels of education. It is therefore anticipated that in the future, more suitable and cost-effective types of AI technology will contribute to an increase in the use of this technology in SE at different levels of education (Heeg & Avraamidou, 2023; Ng et al., 2023).Another result obtained from the research stated that in the issues related to sustainable development goals, emphasis was placed on Quality Education, Good Health and Well Being, Responsible Consumption and Production, and No Poverty. This result has shown that AI has gained a visible level of importance compared to other fields, and that studies focusing on improving the quality of education are more intense. This situation is valuable in showing the need for quality education. The quality of education is one of the most important concepts that positively affects many variables (Al

Husaeni et al., 2024). Akhmadieva et al. (2023) asserted that this situation differs significantly between eastern and western countries. They further observed that the majority of western countries prioritize methodology in studies on AI, whereas eastern countries tend to prioritize educational applications. Moreover, Almasri (2024) asserted that studies tend to prioritize the examination of knowledge and skills. This situation demonstrates that articles on AI place particular emphasis on the cognitive abilities of students, with the objective of enhancing the quality of SE. A further outcome of the research indicated that the most commonly used keywords in articles about AI in SE were AI, science education, computer science education, machine learning, and educational technology. Especially the frequent use of machine learning and ChatGPT among these words can be explained by the fact that ChatGPT takes the lead in AI applications. The fact that the terms ChatGPT and large language models are used in the distribution of words according to years indicates that ChatGPT maintains its superiority even though there are many new tools in the use of AI in education (Cheung et al., 2024; Dunder et al., 2024). A review of the applications' functional areas reveals that, in particular, AI is employed for the purposes of evaluation and feedback. As the range of AI applications continues to expand, it is anticipated that the rate of growth in the utilization of AI in scientific education will also increase (Heeg & Avraamidou, 2023). In addition to the aforementioned applications, AI tools are currently employed in a multitude of domains, including the enhancement of productivity, the examination of data, and the customisation of user experiences. To illustrate, natural language processing (NLP) tools provide efficacious solutions for the analysis of human language, including text analysis and emotion detection (Russell & Norvig, 2016). In particular, highly sophisticated NLP models, such as GPT-3, offer significant convenience in the production of text. Machine learning (ML) is another AI tool that is employed in tasks such as prediction and classification through the process of learning on large datasets. In this field, popular platforms such as scikit-learn and TensorFlow are particularly noteworthy (Brownlee, 2016). The application of machine learning (ML) contributes to the process of decision-making by enhancing the accuracy of predictions in data-intensive sectors such as marketing and healthcare (Chollet, 2021).

Computer vision technologies also facilitate the analysis of visual data and are employed extensively in domains such as object recognition, face recognition and image classification (Kelleher & Tierney, 2018). In particular, within the domain of security, these technologies facilitate the automated analysis of images captured by security cameras. In contrast, voice recognition tools permit users to control devices via voice commands, with applications in numerous aspects of daily life, including the use of Siri or Google Assistant (Goodfellow et al, 2016). Finally, Robotic Process Automation (RPA) represents a significant time-saving measure, particularly within the business sector, through the automation of repetitive processes. These tools, which relieve the burden on the workforce, particularly in areas such as accounting and customer service, are rapidly becoming ubiquitous in business, with platforms such as UiPath and Automation Anywhere (Russell & Norvig, 2016). Upon examination of the most frequently occurring words in the abstracts of the articles, it becomes evident that the primary themes are education, study, AI, learning, paper, and knowledge. This situation has revealed that studies on AI in SE are mostly about knowledge and learning. It is understood that studies expressing high-level thinking skills have not been concentrated yet. In the distribution of these words according to years, the terms evaluation, approach and importance are used more in recent studies. This situation has especially shown that AI is used as an evaluation approach in studies conducted in SE (Haudek & Zhai, 2023). Furthermore, it is evident that the ethical principles pertaining to the utilization of AI are not addressed in the aforementioned studies. This can be attributed to the fact that these studies are still in their nascent stages (Domínguez Hernández & Owen, 2024). Finally, the most recent developments in AI demonstrate that its capabilities are not fully predictable, necessitating the use of limitations and ethical principles in its application. Regarding citations and co-citation analysis, the leading researchers in this field are Cooper, Chin, Dohmen, Zhai, Nehm, Blingsley. These results show that researchers are shifting their fields of study towards AI. In addition, it was determined that the most cited journals included

Journal of Science Education and Technology, Educational Technology & Society, IEEE Transactions on Learning Technologies, Science & Education and Education and Information Technologies. It is understood that these journals, which have an important place in SE, give priority to the subject of AI. These results obtained from the research have shown that it provides important and valuable information in terms of revealing research trends in the use of AI in SE, leading journals and researchers, and countries and institutions that prioritize AI.

It should be noted that the research is not without limitations. First and foremost, the search terms and strategies employed in the research can be regarded as a potential limitation in the process of identifying the included articles. Despite the use of rigorous and comprehensive search terms, it is possible that additional studies may have been included had alternative terms been employed. Moreover, the research was conducted using only the WoS database, which may have resulted in a more limited sample of literature. An examination of different databases might have yielded a different study sample. However, in the process of selecting the sample, particular attention was paid to the inclusion of important indexes and high-quality publications. Consequently, studies included in other indexes were excluded. The inclusion of studies in different indexes may result in a variety of changes to the results. Nevertheless, the validity, reliability, and academic reputation of the sources were rigorously evaluated at each stage of the process. In conclusion, the selected studies were identified through WoS searches at the time the research was conducted. Consequently, studies published subsequent to this date were not included in this research. Therefore, the results obtained are meaningful in line with this information. The findings of the research allow for the formulation of the following recommendations for researchers and future research:

- ❖ Given the dispersion of research efforts across various regions and the variances among countries, it's evident that the United States has spearheaded the bulk of AI-related research. Efforts can be channeled into emphasizing AI within educational programs in other nations.
- ❖ Collaborative international studies can be initiated by partnering with institutions boasting significant publication outputs or by reaching out to researchers affiliated with these institutions.
- ❖ Scholars' ought to scrutinize the contributions of prominent researchers and contemplate submitting their work to reputable journals or seeking publication opportunities therein.
- ❖ Studies on the use of AI in SE focus more on information and evaluation, and studies on different skills and subjects can be emphasized.
- ❖ It is imperative that educators and those aspiring to become educators are made aware of the latest developments in the field of AI. Furthermore, they should be provided with training on the use of different AI tools, and the ethical principles that should be adhered to when using them.
- ❖ A more comprehensive understanding can be achieved by examining studies across a range of indexes.
- ❖ The study revealed a notable increase in the frequency of publications originating from Türkiye in recent times. A review of Türkiye-based studies allows for the examination of trends and comparison with international literature.

References

Abbott, R., & Rothman, E. (2022). *Disrupting creativity: Copyright law in the age of generative artificial intelligence.* Florida Law Review, *75*, 1141. Retrieved from https://heinonline.org/HOL/LandingPage?handle=hein.journals/uflr75&div=34&id=&page=

AI4T. (2024). *Artificial intelligence for and by teachers*. Retrieved from https://www.ai4t.eu/

Akhmadieva, R. S., Udina, N. N., Kosheleva, Y. P., Zhdanov, S. P., Timofeeva, M. O., & Budkevich, R. L. (2023). Artificial Intelligence in Science Education: A Bibliometric Review. *Contemporary Educational Technology*, *15*(4), ep460. https://doi.org/10.30935/cedtech/13587

Al Husaeni, D. F., Al Husaeni, D. N., Nandiyanto, A. B. D., Rokhman, M., Chalim, S., Chano, J., ... & Roestamy, M. (2024). How technology can change educational research? definition, factors for improving quality of education and computational bibliometric analysis. *ASEAN Journal of Science and Engineering*, *4*(2), 127-166. https://doi.org/10.17509/ajse.v4i2.62045

Almasri, F. (2024). Exploring the impact of artificial intelligence in teaching and learning of science: A systematic review of empirical research. *Research in Science Education*, *54,* 1-21. https://doi.org/10.1007/s11165-024-10176-3

Arici, F., Yildirim, P., Caliklar, Ş., & Yilmaz, R. M. (2019). Research trends in the use of augmented reality in science education: Content and bibliometric mapping analysis. *Computers & Education*, *142*, 103647. https://doi.org/10.1016/j.compedu.2019.103647

Berberian, B., Sarrazin, J. C., Le Blaye, P., & Haggard, P. (2012). Automation technology and sense of control: A window on human agency. *PLoS ONE, 7*, e34075. https://doi.org/10.1371/journal.pone.0034075

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). Taxonomy of educational objectives: The classification of educational goals. *Handbook 1: Cognitive domain* (pp. 201–207). New York: McKay.

Broadus, R. N. (1987). Toward a definition of "bibliometrics". *Scientometrics*, *12*, 373-379. https://doi.org/10.1007/BF02016680

Brownlee, J. (2016). *Machine learning mastery with Python: understand your data, create accurate models, and work projects end-to-end*. Machine Learning Mastery.

Büyüköztürk, Ş., Kılıç Çakmak, E., Akgün, Ö., Karadeniz, Ş., & Demirel, F. (2021). *Scientific research methods in education.* Pegem.

Cooper, G., Tang, K. S., & Rappa, N. (2024). Generative artificial intelligence as epistemic authority? Perspectives from higher education. In H. Crompton & D. Burke (Eds.), *Artificial intelligence applications in higher education: Theories, ethics, and case studies for schools*. Routledge (in press).

Chen, X., G. Yu, G, Cheng., & T, Hao. (2019). Research topics, author profiles, and collaboration networks in the top-ranked journal on educational technology over the past 40 years: A bibliometric analysis. *Journal of Computers in Education, 6*(4), 563–585. https://doi.org/10.1007/s40692-019-00149-1

Cheung, K. K. C., Long, Y., Liu, Q., & Chan, H. Y. (2024). Unpacking Epistemic Insights of Artificial Intelligence (AI) in Science Education: A Systematic Review. *Science & Education*, 1-31. https://doi.org/10.1007/s11191-024-00511-5

Chollet, F. (2021). *Deep learning with Python*. Simon and Schuster.

Cope, B., Kalantzis, M., & Searsmith, D. (2021). Artificial intelligence for education: Knowledge and its assessment in AI-enabled learning ecologies. *Educational Philosophy and Theory, 53*(12), 1229–1245. https://doi.org/10.1080/001318572020.1728732

Council of Europe. (2022). *Artificial intelligence and education: A critical view through the lens of human rights, democracy and the rule of law*. Retrieved from https://rm.coe.int/prems-092922-gbr-2517-ai-and-education-txt-16x24-web/1680a956e3

Chichekian, T., & Benteux, B. (2022). The potential of learning with (and not from) artificial intelligence in education. *Frontiers in Artificial Intelligence, 5*, 903051. https://doi.org/10.3389/frai.2022.903051

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of business research*, *133*, 285-296. https://doi.org/10.1016/j.jbusres.2021.04.070

Domínguez Hernández, A., & Owen, R. (2024). 'We have opened a can of worms': Using collaborative ethnography to advance responsible artificial intelligence innovation. *Journal of Responsible Innovation*, *11*(1), 2331655. https://doi.org/10.1080/23299460.2024.2331655

Druga, S., Vu, S. T., Likhith, E., & Qiu, T. (2019). Inclusive AI literacy for kids around the world. In *Proceedings of FabLearn 2019* (pp. 104-111).

Dunder, N., Lundborg, S., Wong, J., & Viberg, O. (2024, March). Kattis vs ChatGPT: Assessment and evaluation of programming tasks in the age of artificial intelligence. In *Proceedings of the 14th Learning Analytics and Knowledge Conference* (pp. 821-827).

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., ... & Wright, R. (2023). Opinion paper: "So what if ChatGPT wrote it?" multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, *71*, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Eker, S., Rovenskaya, E., Langan, S., & Obersteiner, M. (2019). Model validation: A bibliometric analysis of the literature. *Environmental Modelling & Software*, *117*, 43-54. https://doi.org/10.1016/j.envsoft.2019.03.009

Eşiyok, S., & Demircioğlu, M. (2022). OECD ülkelerinin endüstri 4.0 düzeylerinin CRITIC ve CODAS yöntemleri ile değerlendirilmesi [Evaluation of industry 4.0 levels of OECD countries using CRITIC and CODAS methods]. *İstanbul Ticaret Üniversitesi Sosyal Bilimler Dergisi*, *21*(43), 377-398. https://doi.org/10.46928/iticusbe.1076484

García-Peñalvo, F. J., Llorens Largo, F., & Vidal, J. (2023). *The new reality of education in the face of advances in generative artificial intelligence.* https://doi.org/10.5944/ried.27.1.37716

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Haudek, K. C., & Zhai, X. (2023). Examining the Effect of Assessment Construct Characteristics on Machine Learning Scoring of Scientific Argumentation. *International Journal of Artificial Intelligence in Education*, 1-28. https://doi.org/10.1007/s40593-023-00385-8

Heeg, D. M., & Avraamidou, L. (2023). The use of Artificial intelligence in school science: a systematic literature review. *Educational Media International*, *60*(2), 125-150. https://doi.org/10.1080/09523987.2023.2264990

Hudlicka, E. (2016). Virtual affective agents and therapeutic games. In D. D. Luxton (Ed.), *Artificial Intelligence in Behavioral and Mental Health Care* (pp. 81–115). Elsevier.

Jia, F., Sun, D., & Looi, C. K. (2024). Artificial intelligence in science education (2013–2023): Research trends in ten years. *Journal of Science Education and Technology*, *33*(1), 94-117. https://doi.org/10.1007/s10956-023-10077-6

Karagöz, B., & Şeref, İ. (2020) A review of articles on the writing skill: Trends in the web of science database abstract. *Journal of Mother Tongue Education, 8*(1), 67-86. https://doi.org/10.16916/aded.619090

Kelleher, J. D., & Tierney, B. (2018). *Data science*. MIT press.

Kowalczuk-Waledziak, M., Korzeniecka-Bondar, A., Danilewicz, W., & Lauwers, G. (Eds) (2019). *Rethinking teacher education for the 21st century. Trends, challenges and new directions*. Verlag Barbara Budrich. Retrieved from https://library.oapen.org/bitstream/handle/20.500.12657/23733/1006411.pdf?sequence=1&isAllowed=y

Li, Z., & Ironsi, C. S. (2024). The Efficacy of Generative Artificial Intelligence in Developing Science Education Preservice Teachers' Writing Skills: An Experimental Approach. *Journal of Science Education and Technology*, 1-12. https://doi.org/10.1007/s10956-024-10148-2

Malik, A., Khan, M. L., Hussain, K., Qadir, J., & Tarhini, A. (2024). AI in higher education: Unveiling academicians' perspectives on teaching, research, and ethics in the age of ChatGPT. *Interactive Learning Environments*, 1-17. https://doi.org/10.1080/10494820.2024.2409407

Mello, R. F., Freitas, E., Pereira, F. D., Cabral, L., Tedesco, P., & Ramalho, G. (2023). *Education in the age of generative AI: Context and recent developments.* arXiv preprint. https://doi.org/10.48550/arXiv.2309.12332

Miao, F., Holmes, W., Huang, R., & Zhang, H. (2021). *AI and education – Guidance for policy-makers.* United Nations Educational, Scientific and Cultural Organization. Retrieved from https://unesdoc.unesco.org/ark:/48223/pf0000376709

Moher, D., Stewart, L., & Shekelle, P. (2012). Establishing a new journal for systematic review products. *Systematic reviews*, *1*, 1-3. https://doi.org/10.1186/2046-4053-1-1

Mouta, A., Pinto-Llorente, A. M., & Torrecilla-Sánchez, E. M. (2023a). Uncovering blind spots in education ethics: Insights from a systematic literature review on artificial intelligence in education. *International Journal of Artificial Intelligence in Education, 34*, 1166-1205. https://doi.org/10.1007/s40593-023-00384-9

Moral-Muñoz, J. A., Herrera-Viedma, E., Santisteban-Espejo, A., & Cobo, M. J. (2020). Software tools for conducting bibliometric analysis in science: An up-to-date review. *Profesional de la información/Information Professional*, *29*(1), 1-20. https://doi.org/10.3145/epi.2020.ene.03

Mouta, A., Torrecilla-Sánchez, E. M., & Pinto-Llorente, A. M. (2024). Comprehensive professional learning for teacher agency in addressing ethical challenges of AIED: Insights from educational design research. *Education and Information Technologies*, 1-45. https://doi.org/10.1007/s10639-024-12946-y

Mukherjee, D., Lim, W. M., Kumar, S., & Donthu, N. (2022). Guidelines for advancing theory and practice through bibliometric research. *Journal of business research*, *148*, 101-115. https://doi.org/10.1016/j.jbusres.2022.04.042

Mumtaz, S., Carmichael, J., Weiss, M., & Nimon-Peters, A. (2024). Ethical use of artificial intelligence based tools in higher education: Are future business leaders ready?. *Education and Information Technologies*, 1-27. https://doi.org/10.1007/s10639-024-13099-8

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*, 1-11. https://doi.org/10.1016/j.caeai.2021.100041

Ng, D. T. K., Lee, M., Tan, R. J. Y., Hu, X., Downie, J. S., & Chu, S. K. W. (2023). A review of AI teaching and learning from 2000 to 2020. *Education and Information Technologies*, *28*(7), 8445-8501. https://doi.org/10.1007/s10639-022-11491-w

Oh, P. S., & Lee, G. G. (2024). Confronting Imminent Challenges in Humane Epistemic Agency in Science Education: An Interview with ChatGPT. *Science & Education*, 1-27. https://doi.org/10.1007/s11191-024-00515-1

Özbey, D., & Arıcı, F. (2024). Examining the researches on the use of augmented reality technology at primary school level. *Anatolian Journal of Language and Education, 2*(1), 29-46. Retrieved from https://dergipark.org.tr/en/pub/anadoluded/issue/85678/1501045

Page, J., Bain, M., & Mukhlish, F. (2018). *The risks of low level narrow artificial intelligence. In 2018 IEEE international conference on intelligence and safety for robotics* (ISR) (pp. 1-6). IEEE.

Perera, P., & Lankathilaka, M. (2023). Preparing to Revolutionize Education with the Multi-Model GenAI 246. Tool Google Gemini? A Journey towards Effective Policy Making. *Journal of Advances in Education and Philosophy*, *7*(8). 246-253. https://doi.org/10.36348/jaep.2023.v07i08.001

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://doi.org/10.1186/s41039-017-0062-8

Pritchard, A. (1969). Statistical bibliography or bibliometrics. *Journal of documentation*, *25*, 348. Retrieved from https://cir.nii.ac.jp/crid/1570009750342049664

Riera-Negre, L., Hidalgo-Andrade, P., Rosselló, M. R., & Verger, S. (2024, February). Exploring support strategies and training needs for teachers in navigating illness, bereavement, and death-related challenges in the classroom: a scoping review supporting teachers in classroom grief and loss. In *Frontiers in Education* (Vol. 9, p. 1328247). Frontiers Media SA.

Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: a modern approach*. Pearson.

Schlegel, D., & Uenal, Y. (2021). A perceived risk perspective on narrow artificial intelligence. In *PACIS* (p. 44).

Shubhendu, S., & Vijay, J. F. (2013). Applicability of artificial intelligence in different fields of life. Retrieved from https://www.semanticscholar.org/paper/Applicability-of-Artificial-Intelligence-in-Fields-Shubhendu-Vijay/2480a71ef5e5a2b1f4a9217a0432c0c974c6c28c

Smith, L. A., & Petersen, A. C. (2015). Variations on reliability: connecting climate predictions to climate policy. In *Error and uncertainty in scientific practice* (pp. 137-156). Routledge.

Su, J., Zhong, Y., & Ng, D. T. K. (2022). A meta-review of literature on educational approaches for teaching AI at the K-12 levels in the Asia-Pacific region. *Computers and Education: Artificial Intelligence, 3,* 1-18. https://doi.org/10.1016/j.caeai.2022.100065

Tundrea, E. (2020). Artificial Intelligence in Higher Education: Challenges and opportunities. *INTED2020 Proceedings,* 2041–2049. https://doi.org/10.21125/inted.2020.0644

UNESCO. (2023a). *Guidance for Generative AI in Education and Research*. United Nations Educational, Scientific and Cultural Organization. Retrieved from https://www.unesco.org/en/open-access/cc-sa.

UNESCO. (2023b). *Draft AI competency frameworks for teachers and for school students*. Retrieved from https://www.unesco.org/en/digital-education/ai-future-learning/competency-frameworks

Van Eck, N. J., & Waltman, L. (2017). Citation-based clustering of publications using CitNetExplorer and VOSviewer. *Scientometrics*, *111*(2), 1053-1070. https://doi.org/10.1007/s11192-017-2300-7

Van Eck, N., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. S*cientometrics*, *84*(2), 523-538. https://doi.org/10.1007/s11192-009-0146-3

Vučković, A., & Sikimić, V. (2024). Global justice and the use of AI in education: ethical and epistemic aspects. *AI & SOCIETY*, 1-18. https://doi.org/10.1007/s00146-024-02076-x

Wang, P. (2019). On defining artificial intelligence. *Journal of Artificial General Intelligence*, *10*(2), 1-37. https://doi.org/10.2478/jagi-2019-0002

Yusuf, A., Pervin, N., & Román-González, M. (2024). Generative AI and the future of higher education: A threat to academic integrity or reformation? Evidence from multicultural perspectives. *International Journal of Educational Technology in Higher Education*, *21*(1), 1-29. https://doi.org/10.1186/s41239-024-00453-6

Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational research methods*, *18*(3), 429-472. https://doi.org/10.1177/1094428114562629

Article Information Form

Secondary School Students' Disaster Preparedness Perception: A Mixed Method Approach

Ayşenur Kuloğlu[1*] (ID)
Fatma Tutuş[1] (ID)
Merve Özer[1] (ID)

[1] Fırat University, Elazığ, Türkiye, adonder@firat.edu.tr, fakluman@hotmail.com, merve1939@hotmail.com

*Corresponding author

Abstract: The aim of this study was to examine the disaster preparedness perceptions and opinions of Secondary School Students who experienced the most recent major earthquake in Elazığ and Kahramanmaraş in Türkiye. The study population consisted of Secondary School Students in Elazığ and Kahramanmaraş during the 2022-2023 academic year. The sample included 548 participants in the quantitative part selected through non-probabilistic convenient sampling and 40 participants in the qualitative part selected using a criterion sampling method. In this study, a convergent parallel design, a mixed research method, was employed. The "Disaster Preparedness Scale" was used to collect quantitative data. The reliability of the scale was assessed using the Cronbach's Alpha, which was found to be .82. Qualitative data were collected through a semi-structured interview form developed by the researchers. The results indicated that the disaster preparedness perceptions of the participants were at a moderate level. There was no significant difference in the disaster preparedness perceptions among participants based on gender, while significant differences were identified based on city (in favor of students in Elazığ), grade, maternal education level, and paternal education level. Qualitative findings suggested that participants perceived the misinformation in the media after disasters as a significant problem. They also mentioned preparing disaster kits as a precaution after disasters, receiving fire and earthquake-related training at their schools, but feeling that these trainings did not contribute significantly to their preparedness. Participants emphasized the need for disaster awareness education and call for realistic and effective training exercises.

Keywords: Disaster, Disaster Preparedness, Secondary School Students

1.Introduction

The frequency of disasters is increasing globally day by day. Following disasters, there are not only material and spiritual losses but also disruptions in societal life, leading to significant changes (Kolukırık, Arslan & Gökalp Yılmaz, 2022). Many phenomena arising after disasters impact individuals' socio-emotional, physical, cognitive, and psychological conditions, and often lead to lasting harm, especially to children (Garcia, Díaz & Martínez, 2016; Sağlam Şener & Arlıoğlu, 2022). Children exposed to negative living conditions as a result of disasters may encounter severe psychological disturbances. A substantial portion of children experiences post-traumatic stress disorder, depression, and anxiety as a result of disasters, leading to prolonged or enduring psychological distress (Kaya & Özcebe, 2013). In other words, children are significantly affected by both natural disasters (acts of nature) and human-made (intentional or accidental) disasters (Severin & Jacobson, 2020). Considering that disasters affect large areas, various measures can be taken by governments, institutions, and organizations to minimize the material and emotional negative impacts of disasters on people. For instance, providing students with information about disasters, organizing training on how to respond during disasters, and integrating these efforts into government policies can play a significant role in decreasing the effects of disasters (Adanalı, Yiyin & Özenel, 2022). In addition, by implementing necessary security measures in schools (Teyfur, 2021), students can be taught with the understanding that schools are safe places. This can help reduce students' fear of attending school and ensure the continuity of education. Emergency plans for disasters can be expanded, regular training and drills can be conducted, and disaster-resistant structures can be built (Connolly, 2012). According to Karatay and Emini (2022), educational policies can be reviewed and adjusted to ensure that students' educational freedom is not limited and that they

feel secure during disasters. Specifically, adding more disaster education-related achievements to the curriculum can make students more prepared for disasters.

Disaster education is incorporated into various teaching programs within the education system of Türkiye. Subjects such as life sciences, social studies, science and technology, and geography include disaster education outcomes. In particular, the geography curriculum is the program with the most extensive coverage of disaster-related learning outcomes (Başıbüyük & Pala, 2023). In primary and secondary education, topics related to disasters are not allocated a separate subject but are incorporated within the units of existing courses. Topics related to disasters are included in the contents of subjects such as life sciences (grades 1, 2, and 3), science (grades 3, 5, and 8), social studies (grades 4 and 5), and "Our City" (an elective course for grades 5 through 8). In secondary education, these topics are present only in the curriculum of biology (10th grade) and geography (9th, 10th, and 12th grades). It is noted that in primary education, the subject of destructive natural events is discussed as a separate topic, which limits the concept of disasters to solely natural events, and does not conform to modern and contemporary disaster understanding (İnal, Kaya, & Altıntaş, 2018). In addition, the topic of earthquakes in the 4th-grade Social Sciences curriculum is associated with activities such as watching earthquake-related films, earthquake drills, and Earthquake Awareness Week (March 1-7) (Kırıkkaya Buluş, Oğuz Ünver & Çakın, 2011). Starting disaster education in the early years and incorporating relevant knowledge and skills into school curricula are believed to prepare students for disasters and reduce disaster risks (Musacchio et al., 2016). Efforts to develop disaster education curricula and to design them as specialized textbooks, rather than simply integrating them into existing course subjects, can enhance disaster awareness and improve students' perceptions of disaster preparedness (Kırıkkaya Buluş et al., 2011).

Studies have been conducted to measure the disaster awareness levels and preparedness of students at different levels regarding what will happen before, during, and after a disaster (Karakuş & Önger, 2017; Dökmeci & Merinç, 2018; Fathoni, 2018; Ventura & Madrigal, 2020; Şahan & Dinç, 2021; Şekerci, Ayvazoğlu & Çekiç, 2023). In addition, students' preparedness for various potential disasters that could occur at any time and place will enable them to take necessary precautions before a disaster, be aware of the behaviors to adopt or avoid during a disaster, and act consciously after the disaster. As can be seen from the 6.8 magnitude Elazığ earthquake in 2020 and the 7.8 and 7.5 magnitude Kahramanmaraş earthquakes in 2023, Türkiye is an important region at risk of disaster. Therefore, it is important for children and families to learn what to do in case of disaster and how to cope with the difficulties that arise after the disaster. Since students are the most affected individuals during disasters, the governments need to take the necessary measures to minimize the impact on students. This study aims to examine the preparedness levels of students in disaster-prone areas such as Kahramanmaraş and Elazığ, to identify potential shortcomings in their readiness for future disasters, and to shed light on the necessary measures that should be taken. In this sense, the main motivation behind this study was the recognition of the importance of preparing students for disasters and the challenges they may encounter afterward. Thus, thee objective of this study was to examine the disaster preparedness perceptions and views of Secondary School Students who experienced the most recent major earthquake disaster in Elazığ and Kahramanmaraş in Türkiye.

2.Methodology

In this study, a convergent parallel design, a mixed-methods approach, was employed. The mixed-methods design allows for a comprehensive and multidimensional examination of a situation by using both numerical and verbal data concurrently (Gay, Mills, Airasian, 2012; Mills, Gay, 2016). The convergent parallel design is a type of mixed-methods design in which quantitative and qualitative data are collected during the same time frame (Creswell, Plano Clark, 2015). In this design, data are initially analyzed separately and then integrated to draw conclusions.

2.1. Quantitative Study

2.1.1. Participants

The population of the study consisted of Secondary School Students in Elazığ and Kahramanmaraş during the 2022-2023 academic year. A non-probabilistic convenience sampling method was used in sample selection and as a result 548 Secondary School Students from Elazığ and Kahramanmaraş were included in the study. Convenience sampling is a suitable sampling method when participants are easily accessible, data collection is time-efficient. It also facilitates reaching a larger number of participants quantitatively (Büyüköztürk, 2015). The demographic information is provided below.

Table 1

The Participants' Demographic Information

Variables		N	Yüzde (%)
City	Elazığ	228	41,6
	Kahramanmaraş	320	58,4
Gender	Female	301	54,9
	Male	247	45,1
Grade	5. Grade	204	37,2
	6. Grade	271	49,5
	7. Grade	49	8,9
	8. Grade	24	4,4
Maternal Education Status	Primary School	157	28,6
	Secondary School	153	27,9
	High School	156	28,5
	University	82	15,0
Paternal Education Status	Primary School	61	11,1
	Secondary School	120	21,9
	High School	199	36,3
	University	168	30,7
Total		548	100

Of the 548 participants, 228 (41.6%) were enrolled in schools in Elazığ, while 320 (58.4%) attend schools in Kahramanmaraş. In terms of gender, the gender was composed of 301 (54.9%) female students and 247 (45.1%) male students. Among the participants, 204 (37.2%) were in the 5th grade, 271 (49.5%) in the 6th grade, 49 (8.9%) in the 7th grade, and 24 (4.4%) in the 8th grade. Regarding the variable of maternal education, 157 (28.6%) had mothers with primary school education, 153 (27.9%) with middle school education, 156 (28.5%) with high school education, and 82 (15%) with a university degree. In terms of paternal education, 61 (11.1%) participants' fathers completed primary school, 120 (21.9%) completed middle school, 199 (36.3%) completed high school, and 168 (30.7%) had a university degree.

2.1.2. Data collection tool

2.1.2.1. Disaster preparedness scale

The "Disaster Preparedness Scale," developed by Şentuna and Çakı (2020), was used to collect the quantitative data in the study. The scale consists of two sections: demographic variables and scale items. It comprises 13 items, each using a 4-point Likert scale, ranging from "definitely no," "no," "yes," and "definitely yes.". The Kaiser-Meyer-Olkin (KMO) index, used as a criterion for the suitability of the scale items for principal component analysis, was found to have a good fit level at .84 (p = .00). The item-total correlation scores of the scale range between .43 and .63 . The scale is composed of four sub-dimensions: "Disaster Physical Protection," "Disaster Planning," "Disaster Assistance," and "Disaster Warning and Signals." Although there are various debates on the use of the scale, there is an opinion that the use of a 4-point likert-type scale will prevent participants from giving vague answers and encourage them to think more (Garland, 1991). Items in the scale (Definitely yes: 3.21-4.00, Yes: 2.41-3.20, No: 1.61-2.40, Absolutely no: 0.81-1.60) were scored as follows (Şentuna, Çakı, 2020). The lowest score that can be obtained from the whole scale is 4 and the highest score is 42. The reliability of the scale was examined using Cronbach's Alpha which was found to be .82. The CFA of the scale used was examined to see its suitability for secondary school students, and it was seen that this scale could be used for secondary school students. To ensure the scale's reliability for this study, the internal consistency coefficient was calculated and the reliability of the scale was confirmed. It was concluded in the study that the scale showed sufficient fit (p=.00; $\chi 2$/sd=1,99; SRMR= 0,03; RMSEA= 0,04; CFI= .94; TLI= .92).

2.1.2.2. Analysis of quantitative data

The "Disaster Preparedness Scale" and a semi-structured interview form were administered to assess Disaster preparedness perceptions of the participants. The quantitative and qualitative data were analyzed at the end of the data collection process. For the quantitative part of the study, the dependent variable was Disaster preparedness perceptions of the participants, and the independent variables were city, gender, grade, and parental education levels. Data were analyzed using the "IBM SPSS Statistics 22.0" software package. An analysis was conducted to examine whether the data had a normal distribution. It was found that the data exhibited a normal distribution. To be considered as having a normal distribution, skewness and kurtosis values should be between +1.5 and -1.5 (Tabachnick & Fidell, 2013). For the Disaster Preparedness Scale, the skewness value was .226, and the kurtosis value was .642.

For each item on the scale, the arithmetic mean, standard deviation, percentage, and frequency values were calculated. Since the numerical data obtained from the scales were parametric, the t-test was employed for pairwise comparisons, and a one-way analysis of variance (ONEWAY ANOVA) was used for comparisons involving more than two variables.

2.2. Qualitative study

2.2.1. Participants

The study group consisted of 27 female and 13 male participants who were attending schools in Elazığ and Kahramanmaraş during the 2022-2023 academic year. In selecting the study group, criterion sampling method was employed. The criteria for participant selection were that they were in the 8th grade, had experienced a disaster, and had participated in disaster drills. Criterion sampling is a non-probabilistic method in which the sample is composed of individuals, events, or situations that possess the specific qualities defined in relation to the problem (qualifications) under investigation (Büyüköztürk, et al. 2008).

Table 2

The Demographic Variables of the Participants

Variables		N	Percent (%)
City	Elazığ	15	37.5
	Kahramanmaraş	25	62.5
Gender	Female	27	67.5
	Male	13	32.5
Maternal Education Status	Primary School	5	12.5
	Secondary School	9	22.5
	High School	13	32.5
	University	13	32.5
Paternal Education Status	Primary School	5	12.5
	Secondary School	9	22.5
	High School	13	32.5
	University	13	32.5

2.2.2. Data collection tool

2.2.2.1. Semi-structured interview form

To collect qualitative data, a semi-structured interview form was prepared. A semi-structured interview is a flexible interviewing technique in which the researchers have predetermined topic areas, and each participant is asked similar questions (Buran, 2021; Türnüklü, 2000). The interview form consisted of two sections: one section included demographic questions, and the other section contained questions designed to investigate the participants' opinions. To enhance the validity of the data collection tool, feedback was sought from one social studies teacher and two curriculum experts. The prepared interview form was provided to the participants, and their opinions were collected face-to-face on a voluntary basis. The interviews were conducted one-on-one with students in a school setting during suitable time slots over the course of one week. The following questions were posed to the participants regarding their perceptions of disaster preparedness:

1. What are the opinions of Secondary School Students regarding the problems that occur after disasters?
2. What precautions do Secondary School Students take with their families after a disaster?
3. What kind of disaster-related training did Secondary School Students receive at their schools?
4. How did the training they received benefit Secondary School Students during a disaster?
5. What type of disaster-related training would Secondary School Students like to see provided in their schools?
6. According to Secondary School Students, how should disaster drills be conducted?

2.2.2.2. Analysis of qualitative data

The qualitative data were uploaded to the MAXQDA 2018 program and structured using MAXMaps. The data were analyzed using the content analysis method. Content analysis involves transforming words, expressions, and concepts in texts into codes and expressing them in numbers (Kızıltepe, 2021).

2.3. Ethical considerations

Ethics committee permission for this study was obtained from Firat University Social Sciences And Humanities Research Ethics Committee Social And Humanities Research Ethics Committee with the decision dated 13.07.2023 and numbered 2023/13.

3. Findings

This section presents the findings of the analysis conducted to examine the disaster preparedness perceptions of Secondary School Students and their opinions on this topic in terms of various variables.

3.1. Quantitative findings on disaster preparedness perceptions of the participants

3.1.1. Disaster preparedness perceptions of the participants

The arithmetic means and standard deviations were calculated for disaster preparedness perceptions of the participants.

Table 3

Disaster Preparedness Perceptions of the Participants

	N	$\bar{X}$	ss
Disaster Preparedness Perceptions	548	2.66	.45
Disaster Physical Protection	548	2.51	.54
Disaster Planning'	548	2.72	.69
Disaster Assistance	548	3.00	.59
Disaster Warning and Signals	548	2.45	.67

It was found that the participants had a moderate level of disaster preparedness perceptions ($\bar{X}$=2.66). Looking at the sub-dimensions of the disaster preparedness scale in terms of student perceptions, it was found that students had high perceptions in the Disaster Assistance ($\bar{X}$=3.00) sub-dimension, while their perceptions were at a moderate level in the Disaster Physical Protection ($\bar{X}$=2.51), Disaster Planning ($\bar{X}$=2.72), and Disaster Warning and Signals ($\bar{X}$=2.45) sub-dimensions.

3.1.2. Disaster preparedness perceptions of the participants by gender

An independent variables t-test analysis was conducted to examine whether Disaster preparedness perceptions of the participants differed based on their gender.

Table 4

Disaster Preparedness Perceptions of the Participants by Gender

	Gender	N	$\bar{X}$	ss	Sd	t	p.
Disaster Preparedness	Female	301	2.68	.45	546	.920	.35
	Male	247	2.64	.44			

*p.<.05

As shown in Table 4, there was no significant difference in the disaster preparedness perceptions of the participants. However, when examining the means, it was noted that female students ($\bar{X}$=2.68) had higher disaster preparedness perceptions compared to male students ($\bar{X}$=2.64).

3.1.3. Disaster preparedness perceptions of the participants by city

An independent variables t-test analysis was conducted to examine whether the disaster preparedness perceptions of the participants differed based on the city variable.

Table 5

Disaster Preparedness Perceptions of the Participants by City

	City	N	X̄	ss	Sd	t	p.
Disaster Preparedness	Elazığ	228	2.74	.54	546	3.396	.00*
	Kahramanmaraş	320	2.60	.35			

*p<.05

Table 5 revealed a significant difference in the disaster preparedness perceptions of Secondary School Students in Kahramanmaraş in favor of those living in Elazığ (t(546)=3.396; p=.00<.05).

3.1.4. Disaster preparedness perceptions of the participants by grade

A one-way analysis of variance was conducted to examine whether the disaster preparedness perceptions of the participants differed based on their grade.

Table 6

Disaster Preparedness Perceptions of the Participants by Grade

	Grade	N	X̄	sd		Anova					
						Sum of Squares	sd	Mean Squares	F	P	Difference (LSD)
	5.Grade (1)	204	2.71	.47	Between groups	2.517	3	.839	4.188	.006*	
Disaster Preparedness	6. Grade (2)	271	2.67	.45	Within Groups	108.996	544	.200			1>3
	7. Grade (3)	49	2.56	.34	Total	111.514	547				1>4
	8. Grade (4)	24	2.41	.37							2>4
	Total	548	2.66	.45							

*p<.05

As shown in Table 6, a significant difference was found in the disaster preparedness perceptions of the participants in terms of grade (F(3.544) = 4.188; p = .00 < .05). The analysis of the disaster preparedness scale revealed that 5th-grade students had greater disaster preparedness perceptions compared to those in the 7th and 8th grades, and 6th-grade students had higher disaster preparedness perceptions compared to 8th-grade students.

3.1.5. Disaster preparedness perceptions of the participants by maternal education level

A one-way analysis of variance (ANOVA) was conducted to investigate whether the disaster preparedness perceptions of the participants showed significant differences based on maternal education level.

Table 7

Disaster Preparedness Perceptions of the Participants by Maternal Education Level

Maternal Education Level	N	X̄	ss		Anova					
					Sum of Squares	sd	Mean Squares	F	P	Difference
Primary (1)	157	2.61	.48	Between groups	2.513	3	.838	4.180	.006*	
Secondary (2)	153	2.60	.41	Within Groups	109.001	544	.200			4>1
High School (3)	156	2.70	.43	Total	111.514	547				4>2
University (4)	82	2.79	.45							
Total	548	2.66	.45							

*p<.05

As shown in Table 7, there was a significant difference in Disaster preparedness perceptions of the participants in terms of maternal education level (F(3.544) = 4.180; p = .00 < .05). The difference was between the participants whose mothers had university degree ($\bar{X}$=2.79) and those whose mothers were primary school ($\bar{X}$=2.61) or secondary school ($\bar{X}$=2.60) graduates.

3.1.6. Disaster preparedness perceptions of the participants by paternal education level

One-way ANOVA was conducted to examine if there was a significant difference in Disaster preparedness perceptions of the participants concerning paternal education level.

Table 8

Disaster Preparedness Perceptions of the Participants by Paternal Education Level

|	Paternal Education Level	N	$\bar{X}$	sd		Anova					
						Sum of Squares	sd	Mean Squares	F	P	Difference
Disaster Preparedness	Primary (1)	61	2.52	.47	Between groups	3.002	3	1.001	5.017	.002*	
	Secondary (2)	120	2.60	.42	Within Groups	108.512	544	.199			3>1
	High School (3)	199	2.67	.41	Total	111.514	547				4>2
	University (4)	168	2.75	.47							
	Total	548	2.66	.45							

*p<.05

Table 8 revealed a significant difference in disaster preparedness perceptions (F(3,544)=5.017; p=.00<.05). It was found that the participants whose father had completed high school ($\bar{X}$=2.67) exhibited greater disaster preparedness perceptions compared to those whose fathers had finished primary school ($\bar{X}$=2.52). In a similar vein, participants whose fathers were university graduates ($\bar{X}$=2.75) demonstrated higher disaster preparedness perceptions compared to those whose fathers had completed secondary school ($\bar{X}$=2.60).

3.2. Qualitative findings on disaster preparedness of the participants

3.2.1. Problems experienced by the students after disasters

Figure 1

Problems Experienced After Disasters
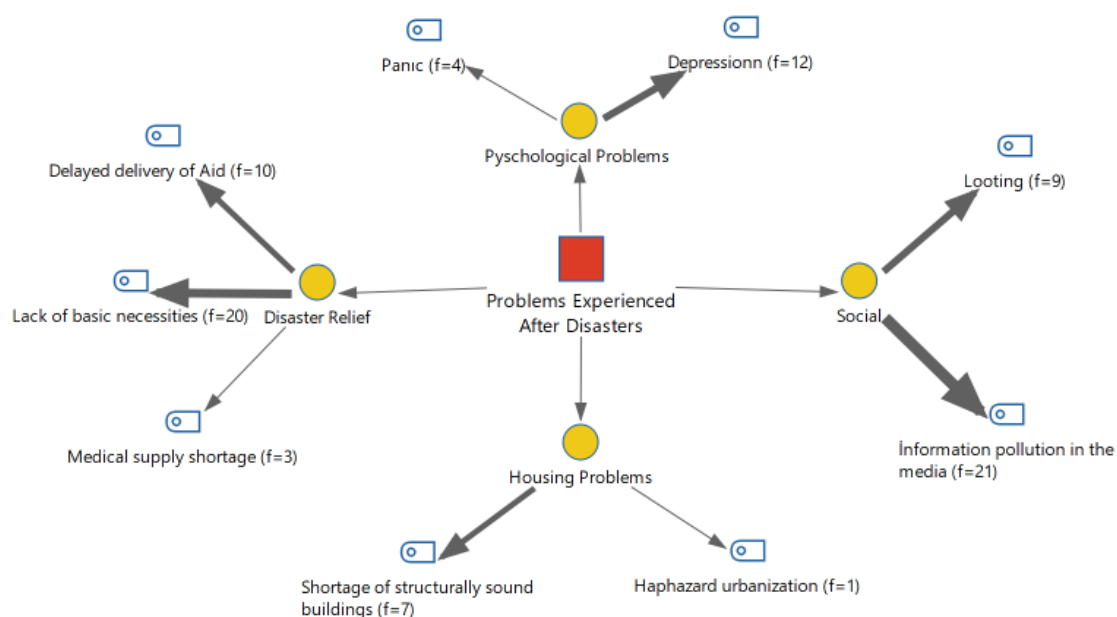
As shown in Figure 1, the participants' views on the problems and deficiencies experienced after disasters were categorized into 4 themes with 9 codes. It was found that the participants primarily focused on information pollution in the media (f=21) and the lack of basic necessities (f=20). The opinions of the participants from Kahramanmaraş revealed that they mainly concentrated on psychological issues and shortages of basic necessities. In contrast, the participants from Elazığ tended to emphasize the delayed response in providing assistance and the information pollution in the media.

S 15: " It left a significant psychological impact, and we felt like we were continuously reliving that moment. It could cause financial effects."

S 18: "During the earthquake we experienced, there were significant deficiencies in the first aid teams. In addition, there were problems with finding shelter, water, and food. There were hygiene problems."

3.2.2. Precautions taken by the participants after disasters

In Figure 2, the opinions of the participants regarding the precautions they took after disasters are presented.

Figure 2

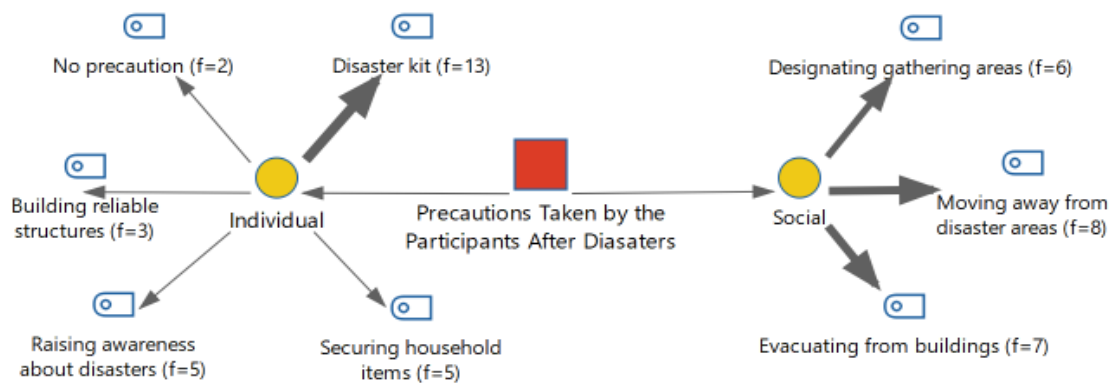*Precautions Taken by the Participants After Disasters*

Figure 2 indicated that the participants' precautions regarding disasters were categorized into two themes and eight codes: individual and collective measures. Students first took measures related to evacuating from buildings (f=7) and moving away from disaster areas (f=8). Subsequently, they prepared disaster kits (f=13) and secured household items (f=5).

S5: "*After the earthquake, we discussed with my family where to stay in the house during the earthquake and where to go if we are outside.*"

S27: "*Right after the earthquake, we moved away from the city, and when we returned home, we prepared a disaster kit and secured the items in the house.*"

3.2.3. Disaster education received by the participants in their schools

Figure 3 presentes the codes regarding the disaster education received by the participants.

Figure 3

Disaster Education Provided to the Participants in Their Schools
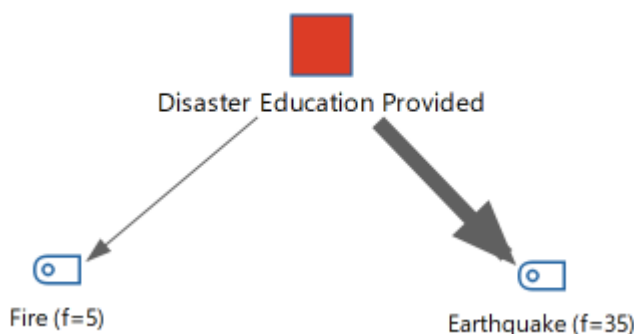
It was observed that students mostly received earthquake-related education.

S40: "An earthquake drill was conducted, and training on actions like drop, cover, and hold was provided."

S1: "We were taught what to do during an earthquake and practiced it. We also received training on how to put out fires during a fire."

3.2.4. Contributions of disaster education received by the participants

In Figure 4, the participants' opinions on the contributions of the disaster education they received regarding disasters were illustrated.

Figure 4

Contributions of Disaster Education Received by the Participants
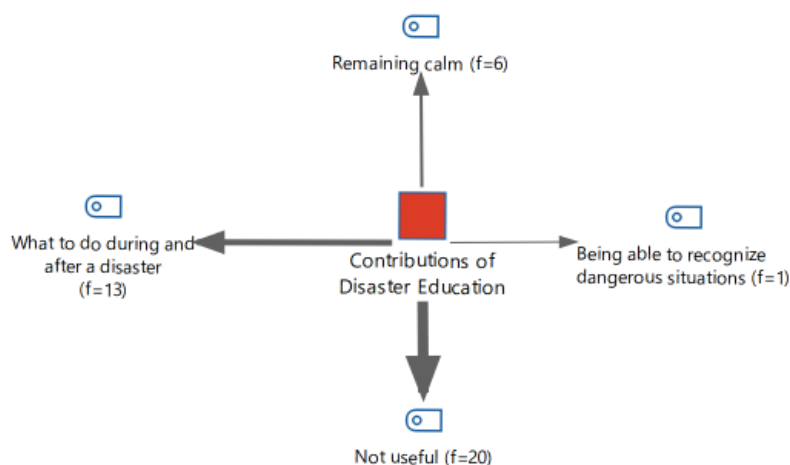

Figure 4 presents the participants' opinions on the contributions of the disaster education they received regarding disasters. While some participants stated that they were informed about what to do during and after a disaster (f=13), others expressed that the education they received was not useful (f=20). Some of the participants' opinions are as follows:

S17: "We received education before the disaster, but it did not help at all. During the disaster, it felt like I had forgotten everything."
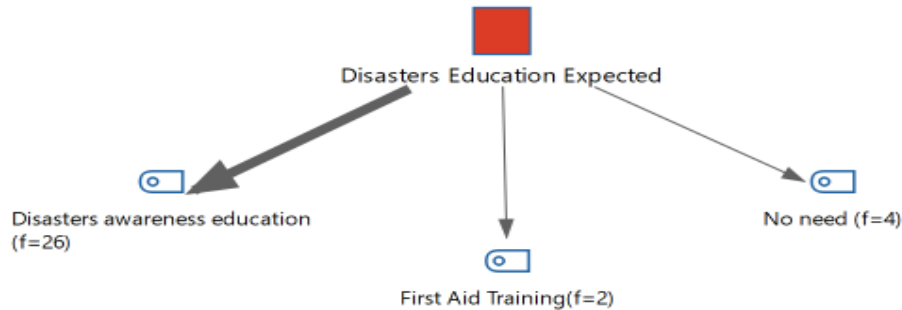
S19: "We had drills at our school before. But I could not stay calm because I was scared of the darkness and earthquakes."

3.2.5. Disaster education the participants expect to receive

Figure 5 shows the disaster education topics that the participants expect to take, represented by codes.

Figure 5

Disasters Education Expected by the Participants

As shown in Figure 5, students had a strong preference for disaster awareness education (f=26). They expressed the necessity of such education to exhibit appropriate behavior during and after disasters. First aid training was considered essential for both disaster situations and daily life.

S2 "Training on how to take necessary measures can be provided. Behaviors like 'Drop, Cover, and Hold On' should not remain purely theoretical, and at least basic first aid training can be provided."
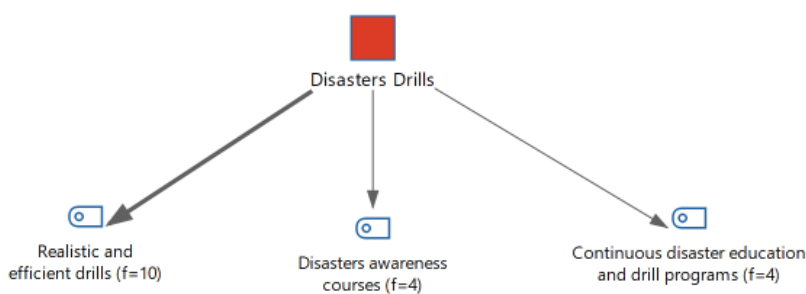
S32 "First aid kits should be provided, and first aid training should be given.""

3.2.6. The participants' recommendations for drills/trainings

Figure 6 presents the recommendations of the participants concerning disaster drills and training, represented using codes.

Figure 6

Recommendations of the Participants for Disaster Drills

The participants emphasized the need for disaster drills and training to be realistic and efficient (f=10). In addition, they expressed a desire for lessons aimed at improving disaster awareness (f=4) and stressed the importance of continuous disaster education and drill programs (f=4).

S16 "I would like to have lessons on how to protect ourselves during disasters and how to act during an earthquake."

S19 "I wish our school would conduct a more realistic drill. We could not adequately prepare for an earthquake because only simple drills have been performed."

4.Results, Discussion and Recommendations

Preparation for disasters, especially earthquakes, involves efforts to minimize the impact of various disasters on individuals. It can be claimed that children are most affected by disasters worldwide. After disasters, children often have difficulties adapting to their education, psychology (Widdyusuf, Muctiarni & Mupita, 2022), and social lives. To minimize these negative effects, it is important to determine children's disaster preparedness perceptions and prepare them for disasters accordingly based on the results. In this context, this study aimed to examine the disaster preparedness perceptions of secondary school students who experienced a major earthquake in Türkiye, specifically in Elazığ and Kahramanmaraş.

The findings indicated that the participants' disaster preparedness perceptions were at a moderate level. In the literature, there are studies which reported both similar and different results (Dereli Toprak & Açıkalın Savaşçı, 2018; Taşkın, 2022). For example, Yıldız, Teeuw, Dickinson, and Robert (2020) investigated earthquake preparedness and risk perception of children, and found that they were generally aware of the earthquake risks in their regions. However, their preparedness levels for such earthquakes were low. Furthermore, they concluded that children who participated in earthquake education programs had higher earthquake awareness, and were able to envisage future earthquakes and possible causes of injury. In contrast, several studies aiming to examine secondary school students' earthquake preparedness reported that the students' earthquake preparedness levels were satisfactory (Alkalash, 2023; Widdyusuf et al., 2022).

It was found in the present study that the participants did not differ by gender. Widdyusuf et. al. (2022) also arrived at a similar conclusion in their study. In contrast to these findings, Cvetković et. al. (2015) found that male students had more disaster-related knowledge compared to female students.

It was also found that there was a significant difference in favor of those in Elazığ regarding the disaster preparedness perceptions of middle school students based on the city variable. A possible reason for this finding could be the significant earthquakes experienced in Elazığ, such as the Sivrice and Elazığ earthquakes. The education received by the children in schools or different social settings, as well as experiencing these earthquakes, may have prepared the children to respond appropriately to disasters. The frequent occurrence of earthquakes in Elazığ, compared to the participants in Kahramanmaraş who may have experienced such a significant disaster for the first time, could be a factor for this finding. Özelmacı (2016) investigated the perceptions of secondary school students regarding disasters and disaster preparedness in Erzurum, Karaman, Kocaeli, and Trabzon. It was concluded that students in Karaman, where disasters are least common in Türkiye, were aware of the impacts of disasters but were not very conscious of preparedness, thinking of a disaster bag as just a simple first aid kit. The lack of experience with disasters could influence children's preparedness perceptions. Cvetković et. al. (2015) studied the knowledge and perceptions of secondary school students in Belgrade about earthquakes as natural disasters, and found that students' knowledge about earthquakes was influenced by personal experiences and the experiences of the head of the family, typically the father. Contrary to these findings, Benzer and Arpalık (2021) did not find a significant difference in the knowledge levels of secondary school students living in different earthquake regions concerning earthquakes based on the city variable.

This study also investigated whether there was a significant difference in disaster preparedness perceptions among secondary school students based on their grade. The findings revealed a significant difference. It was found that participants in 5th-grade had higher disaster preparedness perceptions compared to 7th and 8th graders, and participants in 6th-grade had higher perceptions compared to 8th graders. The reason for this may be that recent disasters were more extensively covered at the primary school level, and it is possible that 5th and 6th-grade students have higher disaster preparedness perceptions compared to students in the 7th and 8th grades contrast, some studies reported that 8th-

grade students had higher knowledge about disasters compared to 5th, 6th, and 7th graders (Adanalı et al., 2022; Benzer & Arpalık, 2021; Piyadeoğlu-Kaya, 2019).

It was also examined whether there was a significant difference between secondary school students' disaster preparedness perceptions based on their parents' education levels. A significant difference between parents' education levels and students' disaster preparedness perceptions was found. A significant difference was observed in participants whose mothers were high school and university graduates, as they had higher disaster preparedness perceptions compared to students whose mothers were elementary and secondary school graduates. Similarly, students whose fathers were high school and university graduates exhibited higher disaster preparedness perceptions than students whose fathers were elementary and middle school graduates. These findings suggest that parents with higher education levels might better inform their children about disasters and that the preparedness of children for disasters is positively correlated with parental education. In contrast, Cvetković et al. (2015) reported that while the education level of parents might not directly affect children's knowledge about earthquakes, the rate at which children of highly educated parents are knowledgeable about earthquakes is statistically higher.

In the qualitative part of the study, the participants were asked about the problems they experienced after the disaster. The responses were mainly about psychological issues, social issues, housing problems, and issues related to disaster aid. The participants in Kahramanmaraş focused particularly on psychological problems and basic needs deficiencies. Tanhan and Mukba (2015) examined the perceptions of secondary school students in Van regarding earthquakes and the psychosocial effects of earthquakes among students. They concluded that earthquakes can cause cognitive, emotional, and behavioral problems in children, as well as psychological issues such as loss of trust and hopelessness about the future. Furthermore, some studies found that students experienced stress disorders after a trauma (Chong-de et al., 2013). Students in Elazığ concentrated on responses related to media misinformation and delays in receiving aid. Koçyiğit (2023) pointed out that the media's coverage encompasses not only communication, aid coordination, and information sharing, but also the generation of false and negative content for specific purposes, which can unnecessarily dominate the media's agenda. The difference in responses may be attributed to the greater destruction in Kahramanmaraş and the lack of basic needs such as food, shelter, and hygiene. The absence of destruction in Elazığ may be related to the different responses given by the students.

In response to the question about the measures they took after disasters, students most commonly reported that they prepared a disaster bag. The lack of basic needs such as food and water, either under the rubble or on the streets, following earthquakes can be attributed to the adoption of this measure. Nevertheless, some participants reported not taking any measures although they experienced a significant disaster. Özkorkmaz et al. also found similar results. The students indicated that they had taken such precautions. Karthikeyan and Rajendran (2020) highlighted the importance of encouraging students to keep water, food, and first aid supplies in their homes or backpacks for an earthquake. They emphasized the importance of creating a disaster preparedness bag. In their study, Özkorkmaz and colleagues (2023) found that approximately 37% of students had prepared a disaster bag.

When asked about the type of disaster training they received at their schools, the participants reported that they received earthquake training and some also received fire training. Türkiye is located in an earthquake-prone region and thus more training shoud be provided. Studies showed that the disaster training provided differs by region. Piyadeoğlu Olcay (2019) examined secondary school students in Gümüşhane and found that the students were educated about disaster awareness for events such as landslides, rockfalls, and floods, which are common in the region.

In terms of the impact of the disaster training they received, the majority of the participants said that it had "no impact." The reason for this finding may be the panic and fear during an earthquake, which may

prevent students from implementing their knowledge. However, the findings of this study contrast with other studies reporting that disaster training can increase disasters awareness (Faupel and Styles, 1993), reduce anxiety after a disaster (Mishra & Suar, 2012), and help children better prepare for disasters (Shiwaku et al., 2007).

In terms of their expectations from disaster training, the majority of the participants stated that disaster awareness education should be provided. Similarly, Özkorkmaz et. al. (2023) emphasized that raising earthquakes awareness of 7th-grade secondary school students was crucial. They argued that earthquake training should be organized, and technological advancements should be included in the programs in order to improve earthquake awareness.

The participants expressed that disaster drills should be more realistic, efficient, and continuous. They also argued that disasters should be taught as a distinct subject. İnal, Kaya, and Altıntaş (2018) highlighted that the disaster education curriculum of primary and secondary education in Türkiye was insufficient.

The study findings indicated that secondary school students had a moderate level of awareness regarding disaster preparedness. However, there is a notable disagreement in the responses of students in terms of earthquake preparedness and preventive measures, as many of them predominantly mentioned the preparation of emergency bags.

In terms of the city variable, the study suggests a more favorable disaster preparedness perception among students in Elazığ. Differences in responses were observed among students residing in Elazığ and Kahramanmaraş concerning the challenges they faced after earthquakes. Specifically, students who experienced the Kahramanmaraş earthquake appeared to undergo more trauma, exhibit higher fears of loss, and demonstrate greater sensitivity to issues related to housing and food shortages.

Research on education revealed positive outcomes for students with parents having a higher level of education. Although the awareness is believed to be linked to education, students expressed that the education received before earthquakes did not significantly contribute to their preparedness. Nevertheless, the students' expressed desire for education in realistic settings and their recognition of the need for continuous education underscore their awareness of the pivotal role of education in disaster preparedness.

4.1 Recommendations

Türkiye is located in an earthquake-prone zone which requires certain measures to be taken. In particular, housing plans should be developed with careful consideration and in collaboration with the government. The number of Disaster and Emergency teams should be increased. Teams capable of providing real life training to students should also be established. Before training children, teachers and parents should receive training in first aid, psychological support, and stress management. In addition, disaster training should be integrated into the curriculum for every course. Furthermore, the drills should be implemented in environments that closely simulate real-world conditions.

References

Adanalı, R., Yiyin, F. T., & Özenel, N. (2022). Disaster awareness of middle school students. *International Journal of Geography and Geography Education (IGGE)*, 47, 56-81. http://dx.doi.org/10.32003/igge.1122725

Alkalash, S. H., Alhashmi Alamer, E. H., Allihyani, A. M., Alhazmi, A. S., Alharthi, R. M., & Bugis, A. M. (2023). Knowledge of and attitude toward disaster preparedness among secondary school students in the western region of Saudi Arabia. *Cureus, 15*(1), 1-13. https://doi.org/10.7759/cureus.33926

Başıbüyük, A., & Pala, S. (2023). Investigation of life science, social studies, and geography curriculum attainments in terms of disaster education. *Erzincan University Journal of Education Faculty, 25*(2), 184-197. https://doi.org/10.17556/erziefd.1063242

Benzer, S., & Arpalık, A. (2021). Knowledge levels of secondary school students living in different earthquake zones about earthquakes. *Journal of Anatolian Cultural Research (JACR), 5*(2), 107-119. https://doi.org/10.15659/ankad.v5i2.138

Buran, A. (2021). Nitel araştırma veri toplama. (Ed. Seggie, F. N., Bayyurt, Y). *Nitel araştırma yöntem, teknik, analiz ve yaklaşımları [Qualitative research methods, techniques, analysis and approaches]*. Anı Yayıncılık.

Büyüköztürk, Ş., Kılıç Çakmak, E., Erkan Akgün, Ö., Karadeniz, Ş., & Demirel, F. (2008). *Bilimsel araştırma yöntemleri [Scientific research methods]*. Pegem Akademi.

Büyüköztürk, Ş. (2015). *Sosyal Bilimler için veri analizi el kitabı [Handbook of data analysis for social sciences]* (21st Ed). Pegem Akademi.

Chong-de, L., Chun, W. X., Yu-di, Z., Zang Wei, Z., Xiao, Z., & Yan, D. (2013). Investigation on mental health state of primary and secondary school students after 30 months of Wenchuan earthquake. *Psychological Development and Education, 29*(6), 631-640. Retrieved from https://caod.oriprobe.com/articles/40876464/Investigation_on_Mental_Health_State_of_Primary_and_Secondary_School_S.htm

Connolly. E. M. (2012). Aligning institutions of higher education emergency preparedness plans with the national response framework. *Journal of Emergency Management, 10*(4), 241-251. https://doi.org/10.5055/JEM.2012.0102

Creswell, J. M., & Plano Clark V. L. (2015). *Karma yöntem araştırmaları tasarımı ve yürütülmesi [Design and conduct of mixed methods research]* (2nd Ed.) (Y. Dede ve S. B. Demir, Çev. Ed.). Anı Yayıncılık.

Cvetković, V. M., Dragičević, S., Petrović, M., Mijalković, S., Jakovljević, V., & Gačić, J. (2015). Knowledge and perception of secondary school students in Belgrade about earthquakes as natural disasters. *Polish Journal of Environmental Studies, 24*(4), 1553- 1561. Retieved from https://rhinosec.fb.bg.ac.rs/bitstream/id/146/274.pdf

Dereli Toprak D., & Açıkalın Savaşçı, F. (2018). Middle school students' conceptual understanding of earthquakes. *SHS Web of Conferences*, *48*, 1-8. https://doi.org/10.1051/shsconf/20184801024

Dökmeci, A. H., & Merinç, F. (2018). Namık Kemal Üniversitesi öğrencilerinin temel afet farkındalığının değerlendirilmesi [Evaluation of basic disaster awareness of Namık Kemal University students]. *Afet ve Risk Dergisi*, *1*(2), 106-113. Retrieved from https://dergipark.org.tr/en/download/article-file/575965

Fathoni, M. (2018). Disaster risk reduction in schools: The relationship of knowledge and attitudes towards preparedness from elementary school students in school-based disaster preparedness

in the Mentawai Islands, Indonesia. *Prehospital and disaster medicine*, *33*(6), 581-586. https://doi.org/10.1017/S1049023X18000778

Faupel, C. E., & Styles, S. P. (1993). Disaster education, household preparedness, and stress responses following Hurricane Hugo. *Environment and Behavior, 25*(2), 228-249. https://doi.org/10.1177/0013916593252004

García, T. C., Díaz, C. V., & Martínez, J. A. C. (2016). Emotional disorders in children victims of natural disasters. *Revista de Enfermería, 39*(6), 18-29. Retrieved from https://europepmc.org/article/med/27548989

Gay, L. R, Mills, G. E., & Airasian, P. (2012). *Educational research: Competencies for analysis and applications* (11th Ed.). Pearson Education.

İnal, E., Kaya, E., & Altıntaş, K. H. (2018). Evaluating the formal education in terms of sufficiency of disaster education in Turkey. A*tatürk Üniversity Journal of Kazım Karabekir Education Faculty, 37*, 114-127. Retrieved from https://dergipark.org.tr/en/pub/ataunikkefd/issue/41724/463823

Karakuş, U., & Önger, S. (2017). 8. sınıf öğrencilerinin doğal afet ve afet eğitimi kavramını anlama düzeyleri [8th grade students' understanding of the concept of natural disaster and disaster education]. *Journal of History Culture and Art Research*, *6*(6), 482-491. http://dx.doi.org/10.7596/taksad.v6i6.1247

Karatay, Z., & Emini, F. T. (2022). Public policy analysis: Primary education policies. *Management and Political Sciences Review, 4*(2), 108-128. https://doi.org/10.57082/mpsr.1184261

Karthikeyan, K., & Rajendran, K. (2020). Teaching Students to Identify Safe and Hazardous Spaces during Earthquakes. *International Journal of Disaster Risk Science, 11*(1), 33-44. https://doi.org/10.1007/s13753-01900246-7

Kaya, E., & Özcebe, H. (2013). Effects of disaster on child's health. *TAF Preventive Medicine Bulletin, 12*(4), 455-464. https://doi.org/10.5455/PMB.1-1340195612

Kırıkkaya Buluş, E., Oğuz Ünver, A., & Çakın, O. (2011). Teachers views on the topic of disaster education at the field on elementary science and technology curriculum. *Necatibey Faculty of Education Electronic Journal of Science and Mathematics Education, 5*(1), 24-42. Retrieved from https://www.academia.edu/download/106572462/EFMED_FBE171.pdf

Kızıltepe, Z. (2021). *İçerik analizi, Nitel araştırma [Content analysis, Qualitative research]* (3th Ed.), (pp. 260-270), Anı Yayıncılık.

Koçyiğit, A. (2023). Combating disinformation and crisis communication in social media in emergency situations: An analysis on the Kahramanmaraş earthquake. *Kastamonu Journal of Communication Research, 10*, 68-86. https://doi.org/10.56676/kiad.1264562.

Kolukırık, S., Arslan, D. A., & Gökalp Yılmaz, G. (2022). Visualization and Social Effects of Forest ires: Media Representations of 2021 Great Antalya-Manavgat Fires. *Journal of Disaster and Risk, 5*(2), 560-580. https://doi.org/10.35341/afet.1086418 Y2

Mills, G. E., & Gay, L. R. (2016). *Educational research: competencies for analysis and applications* (11th Ed.). Pearson Education.

Mishra, S., & Suar, D. (2012). Effects of anxiety, disaster education, and resources on disaster preparedness behavior. *Journal of Applied Social Psychology, 42*(5), 1069-1087. https://doi.org/10.1111/j.1559-1816.2011.00853.x c

Musacchio, G., Falsaperla, S., Bernhardsdóttır, A., Ferreıra, M., Sousa, M., Carvalho, A., & Zonno, G. (2016). Education: Can a bottom-upstrategy help for earthquake disaster prevention? *Bulletin of Earthquake Engineering, 14*, 2069-2086. https://doi.org/10.1007/s10518-015-9779-1

Özelmacı, Ş. (2016). *Examining the perceptions of secondary school students regarding disaster and disaster preparations* (Thesis Number: 450145) [Master Thesis, Gazi University]. Turkish Council of Higher Education Thesis Center.

Özkorkmaz, H., Türk, M. M., Taşyürek, K., Taşyürek, M., Baran, S., & Güler, T. (2023). Examining 7th grade students' awareness about earthquakes. *International Journal of Social and Humanities Sciences Research (JSHSR), 10*(96), 1346-1362. https://doi.org/10.5281/zenodo.8112114

Piyadeoğlu-Kaya O. (2019). *The effect of disaster awareness education given to secondary school students on their knowledge levels: Gümüşhane example* (Thesis Number: 548728) [Master Thesis, Gümüşhane University]. Turkish Council of Higher Education Thesis Center.

Sağlam Şener, T., & Arlıoğlu, H. (2022). The effects of migration as a social disaster on child health. *Journal of Pre-Hospital, 7*(1), 95-107. https://doi.org/10.54409/hod.1069786

Severin, P. N., & Jacobson, P. A. (2020). Types of Disasters. In: Goodhue, C., Blake, N. (eds). *Nursing management of pediatric disaster*. Springer, Cham. https://doi.org/10.1007/978-3-030-43428-1_5

Shiwaku, K., Shaw, R., Chandra Kandel, R., Narayan Shrestha, S., & Mani Dixit, A. (2007). Future perspective of school disaster education in Nepal. *Disaster Prevention and Management, 16*(4), 576-587. https://doi.org/10.1108/09653560710817057

Şahan, C., & Dinç, A. (2021). Afet eğitim merkezlerinde yapılan afet eğitimlerinin ortaokul öğrencileri üzerindeki etkisinin incelenmesi [Examining the effects of disaster training given in disaster training centers on secondary school students.]. *Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi*, (58), 478-500. Retrieved from https://dergipark.org.tr/en/pub/maeuefd/issue/61832/804077

Şekerci, Y. G., Ayvazoğlu, G., & Çekiç, M. (2023). Üniversite öğrencilerinin temel afet bilinci ve farkındalık düzeylerinin saptanması [Determining the basic disaster consciousness and awareness levels of university students]. *Gümüşhane Üniversitesi Sağlık Bilimleri Dergisi*, *12*(1), 74-81. https://doi.org/10.37989/gumussagbil.1136227

Şentuna, B., & Çakı, F. (2020). A scale development study in Balıkesir sampling: Disaster preparedness scale. *İDEALKENT, 11*(31), 1959-1983. https://doi.org/10.31198/idealkent.728896

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Pearson.

Tanhan, F., & Mukba, G. (2015). Investigation of the perception about earthquake based on elementary school students' opinions. *Journal of Human Science, 12*(2), 1581-1601. Retrieved from https://www.j-humansciences.com/ojs/index.php/IJHS/article/view/3360

Taşkın, H. (2022). *The information level and awareness about disasters of 7th grade students in the secondary school: Sapanca case* (Thesis Number: 639984) [Master Thesis, Bezmialem Vakıf University]. Turkish Council of Higher Education Thesis Center.

Teyfur, M. (2021). School security problem in high schools and a model proposal (Izmir province example). *International Social Mentality and Researcher Thinkers Journal, 7*(52), 3067- 3076. http://dx.doi.org/10.31576/smryj.1180

Türnüklü, A. (2000). A qualitative research technique that can be used effectively in educational science research: Interview. *Educational Management in Theory and Practice, 24* (24), 543- 559. Retrieved from https://dergipark.org.tr/en/download/article-file/108517

Ventura, G. L., & Madrigal, D. V. (2020). Awareness and practices on disaster preparedness of students of a public high school in Antique. *Philippine Social Science Journal*, *3*(2), 45-46. Retrieved from https://www.philssj.org/index.php/main/article/download/210/103

Widdyusuf, L., Muktiarni, M., & Mupita, J. (2022). Earthquake disaster preparedness for students of junior high school. *ASEAN Journalof Science and Engineering Education, 2*(2), 129-136. Retrieved from https://ejournal.upi.edu/index.php/AJSEE/article/view/38679

Yıldız, A., Teeuw, R., Dickinson, J., & Roberts J. (2020). Children's earthquake preparedness and risk perception: A comparative study of two cities in Turkey, using a modified PRISM approach. *International Journal of Disaster Risk Reduction, 49*(2), 1-39. https://doi.org/10.1016/j.ijdrr.2020.101666

Article Information Form

Adaptation of Artificial Intelligence Literacy Scale: Latent Profile Analysis

Ali Kırksekiz[1] (ID)
Mehmet Yıldız[1,2] (ID)
Mübin Kıyıcı[3] (ID)
Metin Yıldız[1*] (ID)

[1] Sakarya University of Applied Sciences, Sakarya, Türkiye, alik@subu.edu.tr, mehmetyildiz@subu.edu.tr
[2] Ataturk University, Department of Computer Education and Instructional Technology, Erzurum, Türkiye, yildizz.metin@gmail.com
[3] Sakarya University, Sakarya, Türkiye, mkiyici@sakarya.edu.tr

*Corresponding author

Abstract: Artificial intelligence literacy is vital for individuals' adaptation to the future workforce and societal changes by enabling them to understand and effectively use AI technologies and critically evaluate their impact on society. In this study, the validity and reliability of the artificial intelligence literacy scale in Turkish language were tested and the latent profiles of the students were determined. This methodological study was carried out with a total of 729 students between December 2023 and February 2024. Validity and reliability analyses were conducted with SPSS 27 and AMOS 24, and latent profile analysis was handled with R programming language. According to the results of the CFA analysis of the Artificial Intelligence Literacy Scale, the fit indices were found to be significant (X^2/sd= 3.832, RMSEA=.062, CFI=.949, AGFI=.933, GFI=.960, NFI=.949, TLI=.928, IFI=.916). Considering the Cronbach Alpha value of the scale consisting of 4 sub-dimensions and 12 items, the internal consistency coefficientwas found to be 0.814. Since the lowest BIC value in the latent profile analysis was found in the VVV model, the VVV model was considered as the appropriate one in the study, and the class analyses were carried out through this model. With the LPA analysis, it was designated that the scale was divided into 3 classes. It was determined that the Artificial intelligence literacy scale is a valid and reliable measurement tool. After latent profile analysis, it was found out that the scale was divided into 3 classes.

Keywords: Artificial Intelligence, Literacy, Scale Adaptation, Latent Profile Analysis

1.Introduction

The advancements in computer and internet technologies have introduced a variety of competencies that are essential for individuals, not only in their everyday activities but also to be deemed as skilled workers in the modern labor market. These essential skills are identified as computer literacy, technology literacy, information literacy, internet literacy, and media literacy. However, nowadays, these literacies, which have high levels of interrelationship with each other, are also directly related to digital literacy. Digital literacy broadly refers to the capability to effectively find, organize, comprehend, utilize, communicate, assess, and generate information using digital technologies in a safe and appropriate manner (Law et al., 2018). In terms of its definition, digital literacy can also be regarded as a set of relationships between the knowledge, skills and competencies that individuals should have in responding to the challenges that arise as rapid developments in technology become more effective at every stage of life (McMillan, 2021). Innovations emerging as a result of rapid developments in technology come along with some knowledge, skills and competencies that individuals should have, which leads to the emergence of new literacies.

Artificial Intelligence (AI) is a technology that has ignited significant debate due to its diverse range of products and applications. AI represents the capability of machines to mimic human cognitive functions, including learning, reasoning, and problem-solving, which positions it at the forefront of technological advancement and discussion (Liu et al., 2021; Xu, 2023). AI, which began to come forefront with discussions about digital and human computers at the Paris conference in 1951 (Bruderer, 2016) and was first expressed by John McCarthy (Moloi and Marwala; 2021), has led to revolutionary transformations in many sectors from health to education, finance to law, entertainment to agriculture (Danry et al, 2022; Davenport and Kalakota, 2019; Drach et al., 2023; Minbaleev, 2022; Yin and Moore, 1987; Ruiz-Real et al., 2020). While many people hold favorable opinions about the use of artificial intelligence technologies due to their substantial benefits, it is also necessary to acknowledge potential risks and threats related to information security and ethics (Khawlah et al., 2023). Particularly, the

difficulty of confirming whether a piece of information is produced by artificial intelligence can cause some problems both in the context of information security and ethics.

It may be estimated that AI will transform many labor qualifications in the near future and may even lead to the extinction of many occupational groups, since it has the potential to have a transformative impact in various sectors, including business and human life (Davenport and Kalakota, 2019; Garingan and Pickard, 2021; Puspitaningsih et al., 2022). Hence it can be argued that it is a vital issue, in the context of the competitiveness of individuals, businesses and even countries, to increase the ability of people to produce work with these technologies by including AI technologies in teaching processes in the matter of training qualified labor force. Many scientists carry out scientific research examining the effects of teaching processes enriched with AI technologies on learners in various contexts (Voulgari et al., 2021). With reference to the findings of these studies, it is visible that many countries have developed policies in the context of using AI technologies in teaching processes at all levels starting from early childhood, and developing the knowledge, skills and competencies of learners on how to produce solutions to their own problem situations with these technologies (Puspitaningsih et al., 2022; Williams et al., 2019).

Adapting to the transformation brought about by AI technologies will be possible with individuals possessing high levels of literacy in this field. AI literacy is of great importance for today's digitalized world as it enables individuals to use AI technology to solve problems, encourage analytical, critical and metacognitive thinking, and prepare them for the future (Defeng & Xiaojie, 2020; Puspitaningsih et al., 2022). Additionally, it is crucial to assess the Understanding the AI literacy levels of individuals is crucial. Knowing how a user's proficiency with AI technology affects interactions between humans and AI can help designers create applications that are tailored to the AI literacy levels of their target audience (Wang et al., 2022).

While AI introduces new benefits and opportunities, the biases inherent in these technologies also raise important concerns about ethics and security (Brendel et al., 2021; Wang and Siau, 2019). Individuals need proper training to ensure they use AI responsibly and effectively safeguard their own interests and privacy (Kong et al., 2021). Gaining knowledge, skills, and values related to AI is becoming crucial for individuals. This foundational AI literacy is essential for facilitating effective interactions between humans and machines across social settings, educational spaces, and professional environments (Ali et al., 2019). Within the scope of the study, a latent profile analysis was carried out to determine which policies should be implemented for specific individuals and what types of training should be provided. Latent profile analysis stands out as an effective method to reveal the knowledge levels of individuals on artificial intelligence in the context of different dimensions and to reveal variations.

It is seen that there is a limited number of measurement tools to measure AI literacy in the literature (Laupichler et al. 2023). In Türkiye, there is no measurement tool for determining artificial intelligence literacy for university students. Students' AI literacy can help them understand and manage AI technologies and provide useful information about future job opportunities and career paths. Scale adaptation makes the quantitative measurement of concept of individuals in the current language valid and reliable (Büyüköztürk, et al., 2013). The current study was conducted to address the lack of existing scales in the literature and to perform a latent profile analysis. In the related literature, there are scales developed to determine individuals' AI literacy in various contexts (Ferikoğlu & Akgün, 2022; Hornberger et al., 2023; Hwang et al., 2023; Seong-Won & Lee, 2022). This study aims to bridge this gap by adapting the artificial intelligence literacy scale developed by Wang et al. (2023) into Turkish. It is anticipated that the scale to be obtained can be used as a data collection tool in the needs analysis phase of instructional designs to be realized for all kinds of instructional activities which will be planned to improve the knowledge, and competencies of individuals studying at higher education level on artificial intelligence technologies.

2. Method

In this study, the AI literacy scale was adapted into Turkish and then a latent profile analysis was conducted. The study was carried out between December 2023 and February 2024.

2.1. Population and sample of the research

The study group of the research, which was selected by convenient sampling method, consists of 729 people at Sakarya University of Applied Sciences. In the convenience sampling method, the researcher selects the most accessible sample that offers the greatest savings in time and resources, continuing this approach until a sufficient sample size is achieved (Büyüköztürk, et al., 2013). Due to the fact that the scale is an adaptation study, only Confirmatory factor analysis is considered sufficient (Seçer, 2015). Confirmatory factor analysis is the preferred method for examining the model fit of a scale's factor structure in its original language during the adaptation of a measurement tool developed in another language into Turkish (Seçer, 2015).

Literature reviews reveal varying opinions on the appropriate sample size for scale development and adaptation processes. Bryman and Cramer (2001) advise that the sample size for analysis should be five to ten times the number of items on the scale, while Tabachnick and Fidell (2007) advocate for a minimum of 300 participants, independent of the number of items. In this study, we engaged 729 participants, significantly exceeding the recommended guidelines. This robust sample size, which is at least 20 times the number of items, provides a solid foundation for validating the reliability and accuracy of the research, far surpassing the advised minimum of 300 participants.

2.2. Data collection instruments

2.2.1. Personal information form

This form, which was developed by the researchers, consists of questions investigating the demographic characteristics of the individuals.

2.2.2. Measuring user competence in using artificial intelligence

In this study, the scale developed by Wang et al. (2023) was adapted into Turkish. The aim of the original scale was to develop a valid and reliable scale to measure AI literacy of individuals from different age groups. The original scale consists of 4 sub-dimensions: Awareness, Usage, Evaluation and Ethics. Some of the scale items are;" I can distinguish between smart devices and non-smart devices", "I can skillfully use AI applications or products to help me with my daily work", "I can evaluate the capabilities and limitations of an AI application or product after using it for a while", I always comply with ethical principles when using AI applications or products". The scale is originally a 7-point Likert-type scale consisting of 12 items and 4 sub-dimensions. The scale is scored as Strongly Disagree=1, Disagree=2, Partly Disagree=3, Neutral=4, Partly Agree=5, Agree=6 and Strongly Agree=7. Items 2, Item 5 and Item 11 in the scale are reverse items. The highest score that the person to whom the scale is applied is 84 and the lowest score is 12.

2.2.3. Language equivalence studies of the scale

Scale Language Equivalence Studies are vital to ensure that versions of a scale or questionnaire in different languages function similarly. These studies test whether the concepts that the scale measures in the original language accurately measure the same concepts in the target language. An effective language equivalence study increases the validity and reliability of the scale so that research results are comparable across different cultural and linguistic groups. Therefore, scale language equivalence studies are particularly important in international research and in measurements administered in multilingual communities. In this scale adaptation study, the "process of translation and adaptation of instruments" recommended by WHO was followed (WHO, 2015). Upon securing authorization to use the scale, the original version was independently translated into Turkish by a

bilingual linguist and three academics proficient in English who specialize in the field of artificial intelligence. Then a single text in Turkish was created with the equivalents that best represent each item in the scale. The questions of the Turkish form were checked for semantic integrity by 2 experts who know both source language (English) and target language (Turkish). Finally, the Turkish form was back-translated by a linguist who did not participate in the translation in the first stage and 2 academicians who know both languages. Then the original version was compared with the adapted version by 3 academicians specialized in artificial intelligence.

2.3. Content validity

The content validity rate (CVR) of each item in the scale was calculated institutions (Yeşilyurt, & Çapraz, 2018). The draft scale, which was finalized after the experts' opinions, was applied to 15 students before it was applied to the study sample group. In line with the suggestions received, the scale was evaluated in accordance with the words of the Turkish Language Association in terms of Turkish language and cultural differences. Then 107 students were piloted in the study and the final version of the scale was reached.

2.4. Ethical considerations

Ethics committee permission for this study was obtained from Sakarya University of Applied Sciences Ethics Committee of Rectorate with the decision dated 11.12.2023 and numbered 39, after receiving permission by e-mail from the researchers who developed the original scale. This study was conducted in accordance with the *Principles of the Declaration of Helsinki*.

2.5. Statistical analysis

IBM Statistical Package for Social Science (SPSS) version 27.0 and IBM SPSS Amos version 24.0 were used for data analysis. R programming language was used for latent profile analysis of the scale.

3. Findings

3.1. Structure validity

The path followed in the adaptation of the scale is displayed in Figure 1.

Figure 1

Adaptation Process

Stage	Description
Translation	2 field experts, 2 English language experts, a Turkish language expert
Content and Face Validity	Interviews with 7 field experts, 2 English language experts and a Turkish language expert.
Preliminary Pilot Study	15 students were read aloud and corrections were made in line with the obtained feedback.
Pilot Study	The scale was piloted with 107 students.
Confirmatory Factor Analysis	729 people were applied.
Reliability Analysis	Cronbach Alpha reliability coefficient and split-half reliability coefficient were calculated.
Artificial Intelligence Literacy Scale	As a result of the study, the scale consisting of 12 items and 4 sub-dimensions was validated.

3.1.1. Data analysis

In this context, descriptive statistics of the study group are given in Table 1.

Table 1

Descriptive Statistics for the Sample Group

		Confirmatory Factor Analysis	
		f	%
Gender	Male	422	57.9
	Female	307	42.1
Education Level	Associate Degree	288	39.5
	Bachelor's Degree	385	52.7
	Post-Graduate Degree	57	7.8
Department	Health Sciences	127	17,4
	Computer Programming	115	15,8
	Mechanical Engineering	107	14,7
	Computer Engineering	93	12,8
	Multidimensional Modeling and Animation	49	6,7
	Tourism and Hotel Management	43	5,9
	Civil Engineering	38	5,2
	Mechanical and Metal Technologies	32	4,4
	Mechatronics Engineering	28	3,8
	Welding Technology	27	3,7
	Accounting and Tax Applications	24	3,3
	Electrical and Electronics Engineering	13	1,8
	Other	33	4,5
Total		729	100

Table 1 reveals the statistics for the distribution of the groups according to gender, education level and departments.

The Cronbach Alpha value obtained as a result of the application is presented in Table 2.

Table 2

Cronbach Alpha Value of the Preliminary Application of the Scale

Cronbach Alpha	N	Average	Variance	Standard Deviation	Number of Items
0.959	107	5.275	.667	.81	12

Table 2 shows that the Cronbach Alpha value is 0.959. In this regard, it can be stated that the scale is suitable for analysis (Kline, 2011).

3.1.2. Confirmatory Factor Analysis (CFA)

The path diagram of the CFA is presented in Figure 2.

Figure 2

Confirmatory Factor Analysis Path Diagram
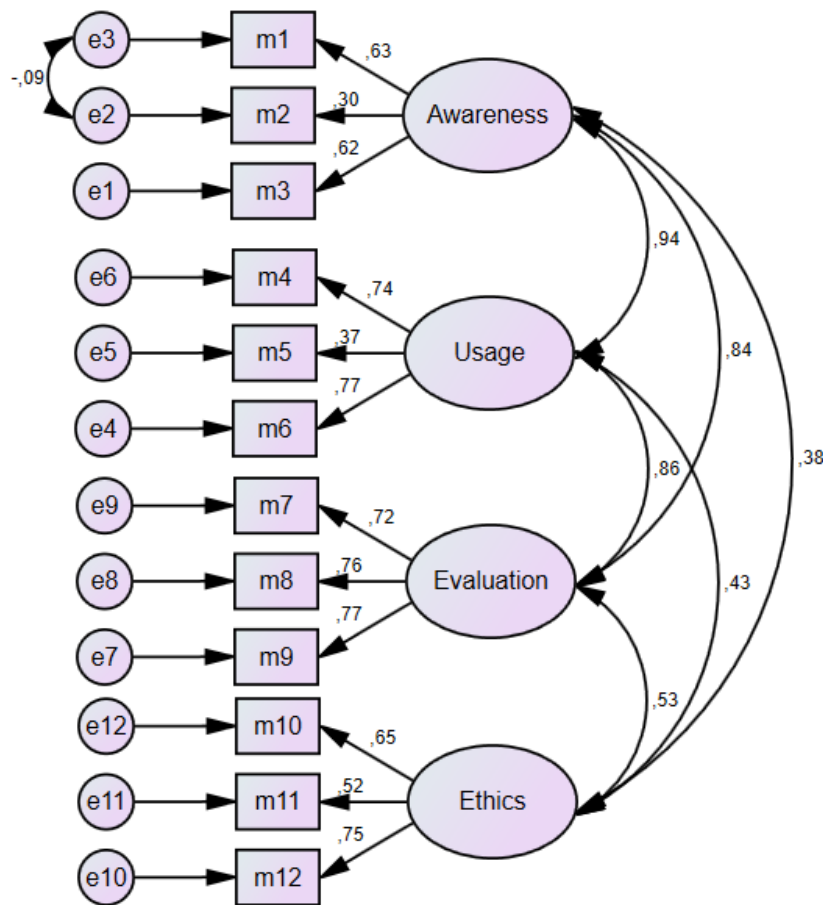
In Figure 2, the factor loadings showing the relationship of each observed variable (m1 - m12) with the related latent construct vary between 0.62 and 0.77. CFA path diagram available in Figure 2, this value is expected to be greater than 0.3, so it was seen that all items met this criterion (Anderson & Gerbing, 1988).

Various goodness-of-fit tests are used to examine the model fit of the scale (McMillan & Schumacher, 2006; Munro 2005; Şimşek 2007; Hooper, Coughlan & Mullen, 2008) are given in Table 3.

Table 3

Confirmatory Factor Analysis Fit Indices

Indices	Reference Value		Measurement	Result
	Acceptable Fit	Good Fit		
CMIN/DF	3< χ2/sd ≤ 5	0< χ2/sd ≤ 3	3.832	Acceptable Fit
RMSEA	.05 ≤ RMSEA ≤ .08	0 ≤ RMSEA ≤ .05	.062	Acceptable Fit
CFI	.90 < CFI≤ .94	.95< CFI≤ 1	.949	Acceptable Fit
AGFI	.85 < AGFI≤ .89	.90< AGFI≤ 1	.933	Excellent Fit
GFI	.85 < GFI≤ .89	.90< GFI≤ 1	.960	Excellent Fit
NFI	.90 < NFI≤ .94	.95< NFI≤ 1	.949	Acceptable Fit
TLI	.90 < TLI≤ .94	.95< TLI≤ 1	.928	Acceptable Fit
IFI	.90 < IFI≤ .94	.95< IFI≤ 1	.916	Acceptable Fit

When Table 3 is examined, it can be concluded that the scale showed a good and acceptable fit.

3.1.3. Correlation coefficients of artificial intelligence literacy scale and sub-dimensions

Correlation coefficients were calculated to determine the relationship between the Artificial Intelligence Literacy Scale (AILS) and the four factors that make up the scale. Table 4 shows that the correlation coefficient between the factors and the whole scale varies between .66 and .81 and the relationship between the sub-dimensions is positive.

Table 4

Correlation Between AILS and Subdimensions

	Awareness	Usage	Evaluation	Ethics
AILS	0.75	0.80	0.81	.66

3.2. Reliability analysis

The results of the calculations are displayed in Table 5.

Table 5

Internal Consistency and Two-Half Reliability Analysis

Test Type		Number of Items	Cronbach Alpha
Literacy Scale		12	0.814
Two-Half Coefficient	1st Section	6	0.706
	2nd Section	6	0.746

Table 5 indicates that the internal consistency coefficient of the scale was 0.814. A Cronbach Alpha coefficient of 0.70 and above is considered appropriate (Büyüköztürk, et al., 2013).

3.3. Latent profile analysis

Response-based analyses aim to reveal structures that are implicit in the data collected from individuals with the help of scales, in other words, the structures that cannot be directly observed. The analysis technique administered in latent classification analysis varies according to the number of measurements, the type of scale on which the measurement is performed, the number of variables measured and whether there is variance between classes. The techniques used in classification analysis are available in Table 6.

Table 6

Summary of Techniques Using Latent Class (Muthén, 2001)

Class	Outcome/ Indicator Scale	Number of Time Points	Number of Outcome / Time Points	Within-Class Variation
LCA	Categorical(u)	Single	Multiple	No
LPA	Categorical(y)	Single	Multiple	No
LCGA	Categorical(u) Categorical(y)	Multiple	Multiple	No
LTA	Categorical(u)	Multiple	Multiple Single	No Yes

Tablo 6 (Continued)

GMM	Continuous(y)	Multiple	Single Multiple	Yes
GGMM	Categorical(u) Continuous(y)	Multiple	Single Multiple	Yes

LCA – latent class analysis, LPA – latent profile analysis, LCGA – latent class growth analysis, LTA – latent transition analysis, GMM - growth mixture modeling GGMM – general growth mixture modeling

Among the techniques summarized in Table 6, especially LCA and LPA analysis are the most common analysis techniques (Ferguson, Moore & Hull; 2020). While LPA deals with continuous cluster indicators, LCA deals with categorical variables (Pastor, Barron, Miller & Davis; 2007). LPA assumes unobserved heterogeneity and classes with specific sub distributions in indicators (Spurk, Hirschi, Wang, Valero & Kauffeld; 2020). Latent class analysis is a statistical method that models the probability of observing specific response patterns within a dataset. This technique helps to identify unobserved, or latent, subgroups within the data based on the responses given (Vermunt, 2022). Steps in LPA/LCA are as follows (Bauer, 2022):

1. Model specification

2. Class enumeration

3. Substantive interpretation of the target model(s)

4. Include predictors and distal outcomes of most likely latent class membership

LPA is less widely used than other latent variable models and, possibly due to that, has long been only available in specialized software packages such as Mplus (Wardenaar, 2021). In the present study, R language was used for LPA and Mclust library (Scrucca, Fraley, Murphy & Raftery; 2023) was used as a library. In addition, Tidylpa library (Rosenberg, van Lissa, Beymer, Anderson, Schell & Schmidt; 2019) was used to calculate fit indices more easily.

In the model evaluation, the VVV model (ellipsoidal, varying volume, shape, and orientation) emerged as the best working model. Akogul & Erisoglu (2017) advise evaluating BIC values in model selection and selecting the model with the lowest BIC value. After deciding on the model, the second step was to determine the best class within the model. In the VVV model, up to 3 classes were analyzed and the results are presented in Table 7.

Table 7

LPA Model Fit Indices Summary

Class	Log Likelihood	AIC	BIC	SABIC	Entropy	LMR Value	LMR p-value	LMR meaning	BLRT Value	BLRT p-value
1	-4196,83	8421,67	8485,97	8441,51						
2	-4076,51	8211,03	8344,23	8252,14	0,78	240,64	0.001	1 < 2	240,64	0,01
3	-3973,91	8035,82	8237,91	8098,2	0,77	205,21	0.001	2 < 3	205,21	**0,01**

Bauer (2022) expresses that the most commonly used fit indices are the Bayesian Information Criterion (BIC), its sample-size-adjusted variant (SABIC), and the Akaike Information Criterion (AIC).

In the study, AIC, BIZ, SABIC, Entropy, LMR and BLRT values were used to decide the number of groups. The lowest values of AIC, BIC, SABIC and CAIC indicate better model fit (Ianculescu, Balog, Cristescu, Iordache & Bajenaru; 2019; Nylund, Asparouhov & Muthén; 2007; Pastor, Barron, Miller & Davis; 2007).

Figure 3

The Density Distribution (a) and Scatter Plot (b)

(a)   (b)
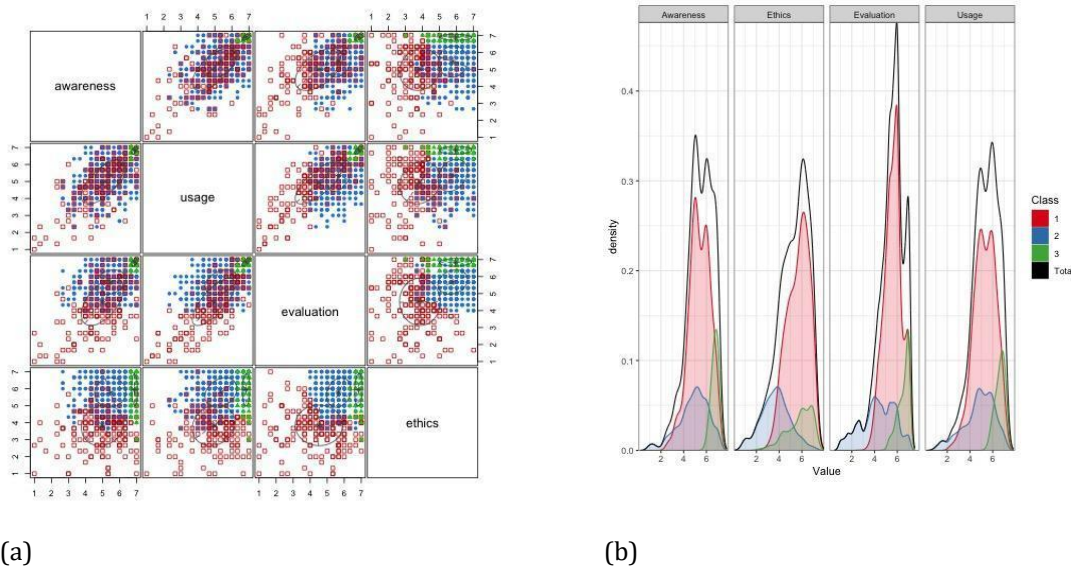
After the completion of the LPA analysis, the density distribution (a) and scatter plot (b) of the classes are displayed in Figure 3.
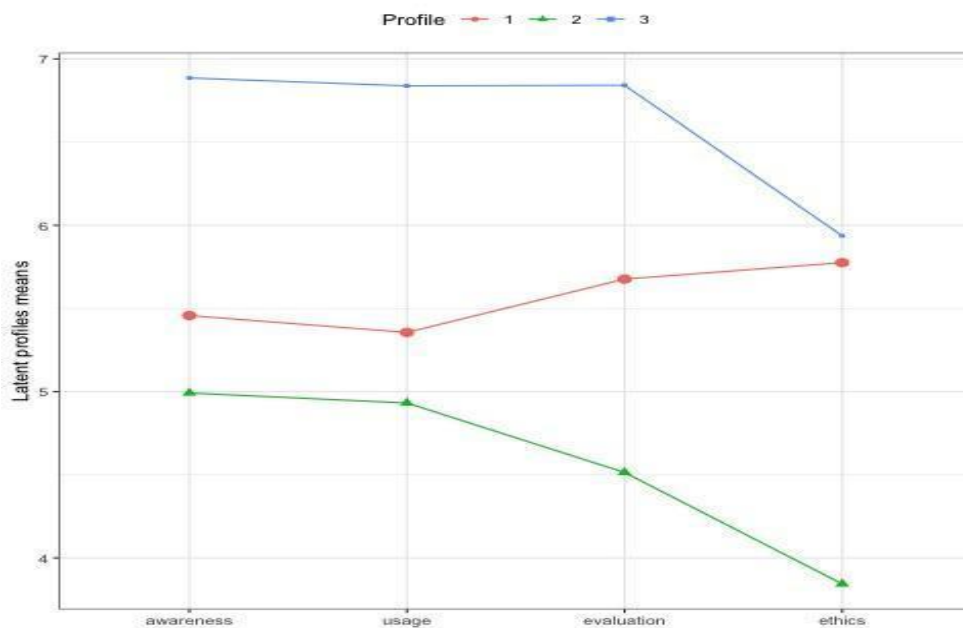
Figure 4

Classes and Means

Figure 3 displays the classes and their distributions as a result of LPA analysis. When Figure 3 is analyzed, the group with the highest average in all 4 indicators is determined as Class 3 and named as high on all aspects. The group with the lowest average in all 4 indicators was determined as Class 2 and named as "low on all aspects". The group with an average mean in all 4 indicators was determined as Class 1 and labeled as "moderate on all aspects".

4. Conclusion and Discussion

The pervasive integration of artificial intelligence across various aspects of life has underscored the importance of AI literacy. The scope of artificial intelligence literacy includes individuals' knowledge, usage, evaluation and ethical usage of artificial intelligence tools. When the literature is reviewed, it is

explicit that there is no measurement tool in Turkish language for university students so as to designate the artificial intelligence literacy levels of individuals. Given the extensive application of artificial intelligence in all areas of life, enhancing AI literacy has become increasingly important. Consequently, the adapted version of this scale is anticipated to make a significant contribution to the academic literature. The objective of this study is to adapt the Artificial Intelligence Literacy Scale (AILS), originally developed by Wang et al. (2022), to assess the AI literacy levels of university students. As part of the adaptation process, the original authors were contacted via email, and permission was secured to translate the scale into Turkish. Then the scale adaptation procedures were started by obtaining ethics committee permission. In the sample distribution of the scale, a special care was given to ensure the necessary heterogeneity by collecting data from different departments. After the language translation and pre-pilot application for adaptation, the pilot study was conducted with 107 participants.

Considering the results, it was uncovered that the Cronbach Alpha value, calculated as the internal consistency coefficient of the study, was 0.81 and this result was consistent with the original scale of 0.83. In the literature, this value of 0.70 and above is accepted as appropriate for the scale to be reliable (Büyüköztürk, et al., 2013; DeVellis 2016). These values indicate that the scale is reliable.

The various fit indices obtained from the adapted scale were as follows: RMSEA=.062, CFI=.949, AGFI=.933, GFI=.960, NFI=.949, TLI=.928, IFI=.916 (Hooper, et al., 2008; Schermelleh-Engel, et al., 2003).

Upon examining the correlation coefficients with the sub-dimensions of the scale, it was observed that they closely mirrored those of the original scale. While there is a high level of positive correlation with each sub-dimension of the scale, respectively: Awareness 0.75, Usage 0.80, Evaluation 0.81 and Ethics 0.66; in the original scale, a high positive relationship was found as follows: Awareness 0.78, Usage 0.72, Evaluation 0.72 and Ethics 0.68.

Latent profile analysis was performed using the Artificial Intelligence Literacy Scale adapted in the study. Three different profiles were observed in Figure 4. Each profile was expressed according to the average of four variables (awareness, usage, evaluation and ethics).

The findings of the study reveal that the participants in the profile designated as Profile 1 and shown with a red line consist of the individuals with medium level values in terms of awareness, usage, evaluation and ethics. These people may represent the individuals who have a generally positive attitude towards artificial intelligence technologies and their usage, and are aware of the potential negative effects of the technology. In a study conducted among graphic design students, it was determined that general attitudes towards AI were positive and this contributed to the creative processes of students (KUM, 2023).

The individuals indicated as Profile 2 in the research and indicated by the green line consist of the individuals with moderate values in terms of awareness and usage, but low values in terms of evaluation and ethics. This finding indicates that the aforementioned individuals understand what AI technologies are and how to use them, but they have deficiencies in evaluating the potential impacts of these technologies and especially in understanding their ethical aspects. These individuals may feel that the benefits of AI technologies outweigh the risks. For members of this profile, an advanced training program focusing on the ethical and evaluative dimensions of AI can be provided. Students' ethical concerns about AI applications are becoming more evident with the increasing use of these technologies in the field of health (Seçer, 2024). Providing transparency and explainability in the decision-making processes of AI is critical to increase users' trust in these systems (Canbay & Demircioğlu, 2021). Moreover, the ability of AI to mimic human intelligence requires in-depth thinking about the ethical responsibilities and consequences of these systems (Ece, 2024).

The group indicated as Profile 3 and shown with the blue line are the individuals with high values in terms of awareness, usage and evaluation but relatively low values in terms of ethics. This profile describes individuals who hold a positive view of AI technologies and their usage, yet remain cautious about the potential adverse impacts these technologies may have. These individuals are willing to take advantage of the benefits of AI technologies but are aware of the potential dangers.

AI can lead to a decrease in the quality of education as students become overly dependent on automated systems to complete their assignment s, thus hindering their intellectual development (Katenova, 2024). The fear of losing one's job due to the capabilities of AI is common among students in various disciplines and many believe that AI can replace traditional educational roles and reduce the human element in teaching (AL-Tkhayneh et al., 2023).

The findings from the Latent Profile Analysis (LPA) indicate that there are significant variations in individuals' AI literacy levels. These differences are likely influenced by factors such as the length and type of individuals' interactions with AI technologies. Last but not least, these findings indicate that policies, research, practices or training to be developed within the scope of AI literacy should be designed for individuals with different profiles.

4.1. Limitations

A limitation of this study is that it does not include participants in the field of social sciences.

References

Akogul, S., & Erisoglu, M. (2017). An approach for determining the number of clusters in a model-based cluster analysis. *Entropy, 19*(9), 452. https://doi.org/10.3390/e19090452

Ali, S., Payne, B. H., Williams, R., Park, H. W., & Breazeal, C. (2019). Constructionism, ethics, and creativity: Developing primary and middle school artificial intelligence education. In *International workshop on education in artificial intelligence K-12 (EDUAI'19)* (pp. 1–4).

AL-Tkhayneh, K., Alghazo, E., & Tahat, D. (2023). The advantages and disadvantages of using artificial intelligence in education. *Journal of Educational and Social Research, 13*(4), 105. https://doi.org/10.36941/jesr-2023-0094

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, 103*(3), 411–423. https://doi.org/10.1037/0033-2909.103.3.411

Bauer, J. (2022). A primer to latent profile and latent class analysis. In *Methods for researching professional learning and development: Challenges, applications and empirical illustrations* (pp. 243-268). Cham: Springer International Publishing.

Brendel, A. B., Mirbabaie, M., Lembcke, T. B., & Hofeditz, L. (2021). Ethical management of artificial intelligence. *Sustainability, 13*(4), 1–18. https://doi.org/10.3390/su13041974.

Bruderer, H. (2016). The Birth of Artificial Intelligence: First Conference on Artificial Intelligence in Paris in 1951?. In: Tatnall, A., Leslie, C. (eds) *International Communities of Invention and Innovation. HC 2016. IFIP Advances in Information and Communication Technology,* vol 491. Springer, Cham. https://doi.org/10.1007/978-3-319-49463-0_12

Bryman, A., & Cramer, D. (2001). *Quantitative data analysis with SPSS release 10 for Windows*. Routledge.

Büyüköztürk, Ş. (2002). Faktör analizi: Temel kavramlar ve ölçek geliştirmede kullanımı [Factor analysis: Basic concepts and use in scale development]. *Kuram ve Uygulamada Eğitim Yönetimi, 32*(32), 470-483. Retrieved from https://dergipark.org.tr/en/pub/kuey/issue/10365/126871

Büyüköztürk, Ş., Çakmak, E., Akgün, Ö., Karadeniz, Ş., & Demirel, F. (2013). *Bilimsel araştırma yöntemleri [Scientific research methods]*. Ankara: Pegem Akademi Yayınları.

Canbay, P., & Demircioğlu, Z. (2021). Endüstri 5.0'a doğru: Zeki otonom sistemlerde etik ve ahlaki sorumluluklar [Towards Industry 5.0: ethics and moral responsibilities in intelligent autonomous systems]. *Ajit-E Online Academic Journal of Information Technology, 12*(45), 106-123. https://doi.org/10.5824/ajite.2021.02.006.x

Cole, D. A. (1987). Utility of confirmatory factor analysis in test validation research. *Journal of Consulting and Clinical Psychology, 55*, 1019-1031. https://psycnet.apa.org/doi/10.1037/0022-006X.55.4.584

Danry, V., Leong, J., Pataranutaporn, P., Tandon, P., Liu, Y., Shilkrot, R., ... & Sra, M. (2022, April). AI-Generated Characters: Putting Deepfakes to Good Use. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts* (pp. 1-5). https://doi.org/10.1145/3491101.3503736

Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future Healthcare Journal, 6*(2), 94-98. https://doi.org/10.7861/futurehosp.6-2-94

Defeng, Q., & Xiaojie, Q. (2020). Curriculum and teaching reform from the perspective of media history. *Philosophy Study, 10*(10). https://doi.org/10.17265/2159-5313/2020.10.005

Drach, I., Petroye, O., Borodiyenko, O., Reheilo, I., Bazeliuk, O., Bazeliuk, N., & Slobodianiuk, O. (2023). The Use of Artificial Intelligence in Higher Education. *International Scientific Journal of Universities and Leadership, 15*, 66-82. https://doi.org/10.31874/2520-6702-2023-15-66-82

Ece, N. (2024). Yapay zeka: Denizcilik sektöründe kullanimi ve swot analizi [Artificial intelligence: its use in the maritime industry and swot analysis]. *Mersin Üniversitesi Denizcilik ve Lojistik Araştırmaları Dergisi, 6*(1), 30-51. https://doi.org/10.54410/denlojad.1491372

Fornell, C., & Larcker, D. F. (1981). Structural equation models with unobservable variables and measurement error: Algebra and statistics. *Journal of Marketing Research, 18*(3), 382-388. https://doi.org/10.1177/002224378101800313

Ferikoğlu, D., & Akgün, E. (2022). An Investigation of Teachers' Artificial Intelligence Awareness: A Scale Development Study. *Malaysian Online Journal of Educational Technology, 10*(3), 215–231. https://doi.org/10.52380/mojet.2022.10.3.407

Ferguson, S. L., Moore, E. W., & Hull, D. M. (2020). Finding latent groups in observed data: A primer on latent profile analysis in Mplus for applied researchers. *International Journal of Behavioral Development, 44*(5), 458-468. https://doi.org/10.1177/0165025419881721

Garingan, D., & Pickard, A. (2021). Artificial intelligence in legal practice: Exploring theoretical frameworks for algorithmic literacy in the legal information profession. *Legal Information Management, 21(*2), 97-117. https://doi.org/10.1017/s1472669621000190

Hooper, D., Coughlan, J., & Mullen, M. (2008). Evaluating model fit: A synthesis of the structural equation modelling literature. In *7th European Conference on research methodology for business and management studies* (pp. 195-200).

Hornberger, M., Bewersdorff, A., & Nerdel, C. (2023). What do university students know about Artificial Intelligence? Development and validation of an AI literacy test. *Computers and Education: Artificial Intelligence, 5*, 100165. https://doi.org/10.1016/j.caeai.2023.100165

Hwang, H. S., Zhu, L. C., & Cui, Q. (2023). Development and Validation of a Digital Literacy Scale in the Artificial Intelligence Era for College Students. *KSII Transactions on Internet and Information Systems (TIIS), 17*(8), 2241-2258. https://doi.org/10.3837/tiis.2023.08.016

Ianculescu, M., Balog, A., Cristescu, I., Iordache, D. D., & Bajenaru, L. (2019). Latent profile analysis in health research: A case study. In *2019 22nd International Conference on Control Systems and Computer Science (CSCS)* (pp. 649-654). IEEE.

Katenova, M. (2024). Artificial intelligence and business school students' performance. *International Journal of Religion, 5*(8), 96-101. https://doi.org/10.61707/6wjvxp71

AL-Tkhayneh, K. M., Al-Tarawneh, H. A., Abulibdeh, E. S. A., & Alomery, M. (2023). Social and Legal Risks of Artificial Intelligence: An Analytical Study. *Academic Journal of Interdisciplinary Studies, 12*(3), 308. https://doi.org/10.36941/ajis-2023-0079

Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling.* New York: The Guilford Press.

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

Laupichler, M. C., Aster, A., Haverkamp, N., & Raupach, T. (2023). Development of the "scale for the assessment of non-experts' AI literacy"–An exploratory factor analysis. *Computers in Human Behavior Reports, 12*, 100338.

Law, N., Woo, D., de la Torre, J., & Wong, G. (2018). A global framework of reference on digital literacy skills for indicator 4.4.2. *UNESCO Institute for Statistics.* Retrieved from https://unesdoc.unesco.org/ark:/48223/pf0000265403

Liu, N., Shapira, P., & Yue, X. (2021). Tracking developments in artificial intelligence research: Constructing and applying a new search strategy. *Scientometrics, 126*(4), 3153-3192. https://doi.org/10.1007/s11192-021-03868-4

McMillan, C. T. (2021). *Posthumanism in digital culture: Cyborgs, Gods and Fandom.* Emerald Publishing Limited.

McMillan, J. H., & Schumacher, S. (2006). *Research in education: Evidence-based inquiry.* Pearson.

Minbaleev, A. (2022). The concept of "artificial intelligence" in law. *Bulletin of Udmurt University Series Economics and Law, 32*(6), 1094-1099. https://doi.org/10.35634/2412-9593-2022-32- 6-1094-1099

Moloi, T., & Marwala, T. (2021). A High-Level Overview of Artificial Intelligence: Historical Overview and Emerging Developments. In *Artificial Intelligence and the Changing Nature of Corporations.* Springer, Cham. https://doi.org/10.1007/978-3-030-76313-8_2

Munro, B. H. (2005). *Statistical methods for health care research.* Lippincott Williams & Wilkins.

Muthén, B. O. (2001). Latent variable mixture modeling. In *New developments and techniques in structural equation modeling* (pp. 21-54). Psychology Press.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(4), 535-569. https://doi.org/10.1080/10705510701575396

Pastor, D. A., Barron, K. E., Miller, B. J., & Davis, S. L. (2007). A latent profile analysis of college students' achievement goal orientation. *Contemporary Educational Psychology, 32*(1), 8-47. https://doi.org/10.1016/j.cedpsych.2006.10.003

Puspitaningsih, S., Irhadtanto, B., & Puspananda, D. (2022). The role of artificial intelligence in children's education for a digital future. *Kne Social Sciences, 5th International Conference on Education and Social Science Research (ICESRE),* 642-647. https://doi.org/10.18502/kss.v7i19.12483

Rosenberg, J. M., van Lissa, C. J., Beymer, P. N., Anderson, D. J., Schell, M. J., & Schmidt, J. A. (2019). *tidyLPA: Easily carry out latent profile analysis (LPA) using open-source or commercial software* [R package]. Retrieved from https://data-edu.github.io/tidyLPA/

Ruiz-Real, J., Uribe-Toril, J., Arriaza, J., & Valenciano, J. (2020). A look at the past, present and future research trends of artificial intelligence in agriculture. *Agronomy, 10*(11), 1839. https://doi.org/10.3390/agronomy10111839

Scrucca, L., Fraley, C., Murphy, T. B., & Raftery, A. E. (2023). *Model-based clustering, classification, and density estimation using mclust in R.* Chapman and Hall/CRC.

Seçer, E. (2024). Technostress levels of physiotherapy and rehabilitation students, related factors and awareness of the use of artificial intelligence in health: A cross-sectional study. *Turkiye Klinikleri Journal of Health Sciences, 9*(1), 127-136. https://doi.org/10.5336/healthsci.2023-100746

Seçer, İ. (2015). *Psikolojik test geliştirme ve uyarlama süreci: SPSS ve LISREL uygulamaları [Psychological test development and adaptation process: SPSS and LISREL applications].* Anı Yayıncılık.

Seong-Won Kim, Youngjun Lee. (2022). The artificial intelligence literacy scale for middle school students. 국컴퓨터정보학회논문지, *27*(3), 225-238. https://doi.org/10.9708/jksci.2022.27.03.225

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8*(2), 23-74. Retrieved from https://www.stats.ox.ac.uk/~snijders/mpr_Schermelleh.pdf

Spurk, D., Hirschi, A., Wang, M., Valero, D., & Kauffeld, S. (2020). Latent profile analysis: A review and "how to" guide of its application within vocational behavior research. *Journal of Vocational Behavior, 120*, 103445. https://doi.org/10.1016/j.jvb.2020.103445

Şimşek, Ö. F. (2007). *Yapısal eşitlik modellemesine giriş: Temel ilkeler ve LISREL uygulamaları [Introduction to structural equation modeling: basic principles and LISREL applications.].* Ekinoks.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics.* Allyn & Bacon.

Vermunt, J. K. (2022). Latent class analysis. In *International Encyclopedia of Education*, Fourth Edition. Oxford: Elsevier.

Voulgari, I., Zammit, M., Stouraitis, E., Liapis, A., & Yannakakis, G. (2021, June). Learn to machine learn: Designing a game based approach for teaching machine learning to primary and secondary education students. In *Proceedings of the 20th Annual ACM Interaction Design and Children Conference* (pp. 593-598). https://doi.org/10.1145/3459990.3465176

Wang, B., Rau, P., & Yuan, T. (2023). Measuring user competence in using artificial intelligence: Validity and reliability of artificial intelligence literacy scale. *Behaviour & Information Technology, 42*(9), 1324–1337. https://doi.org/10.1080/0144929x.2022.2072768

Wang, W., & Siau, K. (2019). Artificial intelligence, machine learning, automation, robotics, future of work and future of humanity: A review and research agenda. *Journal of Database Management, 30*(1), 61–79. https://doi.org/10.4018/JDM.2019010104

Wardenaar, K. J. (2021). *Latent profile analysis in R: A tutorial and comparison to Mplus. Retrieved* from https://psyarxiv.com/wzftr/download

WHO (2015). *Process of translation and adaptation of instruments.* Retrieved from http://www.who.int/substance_abuse/research_tools/translation/en/

Williams, R., Park, H., Oh, L., & Breazeal, C. (2019). Popbots: Designing an artificial intelligence curriculum for early childhood education. *Proceedings of the AAAI Conference on Artificial Intelligence, 33(*01), 9729-9736. https://doi.org/10.1609/aaai.v33i01.33019729

Xu, F. (2023). The link between artificial intelligence and computer technology. *Sixth International Conference on Intelligent Computing, Communication, and Devices (ICCD 2023).* https://doi.org/10.1117/12.2683091

Yin, R. K., & Moore, G. B. (1987). The use of advanced technologies in special education: Prospects from robotics, artificial intelligence, and computer simulation. *Journal of Learning Disabilities, 20*(1), 60-63. https://doi.org/10.1177/002221948702000111

Yeşilyurt, S., & Çapraz, C. (2018). Ölçek geliştirme çalışmalarında kullanılan kapsam geçerliği için bir yol haritası [A roadmap for content validity used in scale development studies]. *Erzincan Üniversitesi Eğitim Fakültesi Dergisi, 20*(1), 251-264. https://doi.org/10.17556/erziefd.297741

Article Information Form

Authors Notes: Authors would like to express their sincere thanks to the editor and the anonymous reviewers for their helpful comments and suggestions.

Authors Contributions: All authors contributed equally to the study.

Conflict of Interest Disclosure: No potential conflict of interest was declared by authors.

Copyright Statement: Authors own the copyright of their work published in the journal and their work is published under the CC BY-NC 4.0 license.

Supporting/Supporting Organizations: No grants were received from any public, private or non-profit organizations for this research.

Ethical Approval: Ethics committee permission for this study was obtained from Sakarya University of Applied Sciences Ethics Committee of Rectorate with the decision dated 11.12.2023 and numbered 39.

Plagiarism Statement: This article has been scanned by iThenticate.