TOJET
The Turkish Online Journal of Educational Technology

ISSN 2146-7242

Turkish Online Journal of Educational Technology

Volume 25, Issue 2
April 2026

TOJET
01.04.2026

THE TURKISH ONLINE JOURNAL OF EDUCATIONAL TECHNOLOGY

April 2026
Volume 25 – Issue 2

Prof. Dr. Aytekin İşman
Editor-in-Chief

ISSN: 2146 - 7242

Indexed by
Education Research Index
ERIC
EBSCOhost – Current Abstracts
EBSCOhost – Education Research Index
EBSCOhost – TOC Premier
Cabell's Directories
Index Copernicus Journal Master List

Published in TURKEY

Contact Address:
Prof. Dr. Aytekin İŞMAN
TOJET, Editor in Chief
Sakarya-Turkey

Message from the Editor-in-Chief

Editors' Note

The April 2026 issue of *The Turkish Online Journal of Educational Technology (TOJET)* brings together **twelve rigorously selected studies authored by an international body of scholars**, advancing the field through theoretical depth, methodological diversity, and a strong emphasis on the evolving relationship between technology, pedagogy, and society.

A defining characteristic of this issue is its critical engagement with **artificial intelligence in education**. The study by **Ahmet Adalıer et al.** demonstrates, through explainable AI models, that student achievement emerges from a multidimensional interaction of psychological, behavioral, and environmental variables. Complementing this, **Ali Akçın et al.** explore teachers' perceptions of generative AI, highlighting both its pedagogical affordances and ethical tensions.

From a digital culture perspective, **Bilge Yüce Abay & Aytekin İşman** examine social media use through motivational frameworks, while **Burak Hüseyin Günsel et al.** emphasize the role of organizational unlearning in sustaining institutional quality and innovation.

The issue further positions **AI literacy as a core educational competency**. **Aysan Kolahdouzipour & Ayşegül Kulavuz-Onal** investigate AI-supported academic writing, demonstrating the importance of critical engagement, while **Nurcan İnan et al.** develop and validate a prompt-writing rubric that operationalizes effective human–AI interaction skills.

Innovative pedagogical practices are also foregrounded. **Dr. Boggs** shows that student-generated video production enhances higher-order thinking and digital literacy, whereas **Taleh Mirzayev** provides a comparative analysis of authentic and transformational leadership, linking leadership styles to teacher well-being and institutional effectiveness.

In the domain of inclusive and specialized education, **Sabit Menteşe** presents a bibliometric analysis of twice-exceptionality research, identifying global trends and research gaps, while **Tülay Ekici** highlights the multidimensional nature of music teacher training, integrating cognitive, technical, and affective competencies.

Extending the discussion to structural dimensions of education, **Yuki Amaki** examines the concept of "Oya Gacha" through cultural reproduction theory, demonstrating how students interpret educational inequality across cultural contexts.

Finally, Yogaranee Sakthivel contributes to the issue by offering a contextually grounded and empirically robust analysis of contemporary educational dynamics, providing critical insights into how emerging socio-cultural and pedagogical variables reshape learning processes, learner engagement, and equity in diverse educational environments.

Taken together, the twelve articles in this issue collectively signal a paradigm shift from technology-centered discourse toward a **human-centered, ethically grounded, and system-aware understanding of educational technology**. Rather than positioning technology as an end, these studies emphasize its role as a transformative and context-dependent mediator within complex educational ecosystems.

As TOJET continues to position itself at the forefront of global scholarly dialogue, this issue not only reflects current trends but also actively shapes the future research agenda by bridging disciplines, methodologies, and paradigms. It calls for a redefinition of educational technology—not as a set of tools, but as a complex, adaptive, and value-laden ecosystem.

We extend our sincere appreciation to all authors, reviewers, and contributors whose work continues to elevate the scientific rigor and global impact of TOJET.

Call for Papers:
TOJET welcomes academic studies in the field of educational technology. Submitted articles may address topics such as the use of technology in classrooms, the impact of technology on learning, and the perspectives of

students, teachers, administrators, and the community on educational technology. Such studies will enhance the quality of theoretical and practical approaches in educational technology.

Article Submission Criteria:

- Submitted articles must be original, unpublished, and not under consideration by another publication.
- Articles may cover a wide range of topics, including assessment, attitudes and beliefs, curriculum design, equity, applied research, learning theories, sociocultural issues, and educational practices for special populations.

Warm regards,

Prof. Dr. Aytekin İşman
Editor-in-Chief, TOJET

Editor
Prof. Dr. Teresa Franklin
Ohio University
April 2026

Editorial Board

Table of Contents

A Study on Improving the Singing Skills of Prospective Music Teachers

Tülay EKİCİ
Dokuz Eylül University, Buca Faculty of Education, Department of Music Education
ORCID: 0000-0001-9587-7849
tulay.ekici@deu.edu.tr

ABSTRACT
The human voice, the oldest, most natural and most fundamental musical instrument known, serves as a primary channel for everyday verbal communication and through the art of singing, transforms into an aesthetic and expressive mean of musical communication. Therefore, the optimal utilization of an individual's vocal potential, the use of the voice in its natural state, and the maintenance of vocal health constitute critical and sensitive issues. Furthermore, the human voice may be regarded as one of the most fundamental and natural instruments in the processes of music learning and teaching, while the song form can be considered among the most basic forms of musical expression. Considering that singing is the goal and natural outcome of voice training, developing the singing skills of prospective music teachers will contribute to their greater confidence and success both during their undergraduate education and in their future professional lives. In this context, the purpose of this research is to evaluate the importance of developing the singing skills of prospective music teachers and to offer recommendations. The research conducted a literature review and employed the interview technique, a qualitative research method. The study group consisted of prospective music teachers studying in the Department of Music Education at Dokuz Eylül University's Buca Faculty of Education and taking a vocal training course. At the end of the research, suggestions were presented in line with the findings obtained for the development of singing skills of music teacher candidates.
Keywords: Prospective Music Teacher, Voice Education, Song, Singing Skill.

INTRODUCTION

The art of song

The highest expression of art is music, and the highest expression of music is the art of singing. Music lies at the foundation of universal existence and contributes to our understanding of the harmony, rhythm, and aesthetics necessary for the establishment of order. At the core of the universe lies sound, rhythm, and tone that is, music in its entirety. The human being constitutes the most important musical instrument in the practice of music (Khan, 1994). Consequently, the highest expression of music is manifested in the art of singing. Since emotion represents the fundamental element of art, singing serves as the most direct and profound medium through which emotional expression can be conveyed. The human voice is the only instrument that is a part of the human body and connected to the human soul. The true impulse in the art of singing is the reflection of the emotions belonging to the human soul (Rogers, 2013).

It is widely accepted that music began with singing, and that singing is the most universal musical activity (Potter and Sorrell, 2014). It is even thought that singing existed before musical instruments, before writing and notation, and may even have existed before speech. Although little is known about its origins, it seems plausible that one of the early functions of singing was oral communication (Bicknell, 2015).

Art is the performance or accomplishment of a task based on skill and understanding. In this context, the art of singing refers to the skillful control, manipulation, and use of the voice in music, based on both technique and comprehension. As a musical ability, it can be appreciated both as an end in itself and as a means to an end. As a means to an end, the art of singing represents a form of understanding that involves various intellectual interactions, such as acquiring information, analyzing, and evaluating. As an end in itself, it encompasses multiple levels of musical experience, including emotional, perceptual, and creative dimensions (Rao, 1988).

The art of singing, defined as the special use of the human voice in its biological, psychological, emotional, and aesthetic aspects (Rao, 1988), is also described, with a similar approach, as a complete body and mind experience where physical, intellectual, emotional, and creative qualities coexist (Morrison, 2006). Davran (1997) offers a beautiful definition of the art of singing that encompasses all three fundamental dimensions of human existence: mind, body, and soul (emotion): "The art of singing is the act of gaining control over all the muscles in our body involved in the act of singing, using these muscles as needed, and thus achieving access to the secrets of the human body and soul."

The act of singing

The act of singing as an artistic practice is directly related to the musical structure itself. A singer must be able to control their voice, make the necessary changes and improvements to their voice to meet the demands of the music, and evaluate the results of their actions. The ability to make changes and improvements to the voice depends on the singer's skill and understanding, and is guided by the aesthetic requirements of the music. Therefore, the art of singing can be considered a definable system of artistic processes. As a matter of fact, the act of singing as an art is the ability to use the body in a highly diverse and skillful way for expressive purposes. Thus, like all artistic skills, specific singing skills are practiced with the "standards, values, ideals, and goals specific to the art form" in mind. We practice these skills in light of the best knowledge we possess. When used for musical purposes, singing can be defined as a rational act (Sparshott, 1982). The artistic processes of singing can be described as a three-part process including:

• The artist's intention (what the singer plans to do to produce singing tone and make vocal enhancements),
• The processes initiated (the movements used to transform the voice as a sound generator into sound as a musical instrument), and
• The results obtained.

According to Rao (1988), the art of singing involves the deliberate transformation of the voice to create music, “the intelligent modification, control, and manipulation of the voice based on skill and understanding.” More specifically, the art of singing is concerned with how muscles move to create musical effects through the conscious manipulation of many parts of the body, especially the muscles used in breathing, phonation, resonance, and articulation. This deliberate movement of muscles is necessary to modify the voice in the context of music, that is, to change the quality and duration of the pitch. While muscle movements in speech are reflexive and passive, in singing, the movements of the vocal muscles are necessarily deliberate and active. Therefore, even muscle movements that are involuntary in singing and more dependent on the large muscle movements of breathing and articulation become part of the directed, rational actions initiated to produce and control the pitch. In short, to be able to produce music with the human voice, all parts of the vocal system must function as an artistic whole.

McKinney (1994) states: “Being able to manipulate sounds means you know the nature of sound, you know how musical instruments work, and you understand the relationship between the sound instrument and the physical processes that govern it. Being able to manipulate sounds also means you set standards for good sounds, especially by listening to renowned artists who have achieved lasting success and continue to demonstrate vocal freedom and tonal beauty.”

Singing education

As O’Bryan and Harrison (2014) note, singing is a nearly universal human practice, appearing across cultures for a wide array of purposes and in diverse contexts. People sing for joy, love, enlightenment, or entertainment, as well as for sorrow, anger, or emotional and spiritual support—expressing the human spirit through music, either individually or collectively. For many singers, formal training becomes an essential pursuit, driven by the desire to convey these emotions as purely and flawlessly as possible. As singing evolved into a more public, professional, and respected art form, the demand for structured training increased to meet higher standards of artistic excellence. While some individuals are naturally gifted in singing, the skill itself is, contrary to popular belief, complex and requires dedicated mastery. This complexity arises from the need for precise coordination and conditioning of the entire body, making training essential for the development of singing abilities. Although contemporary pedagogical approaches have undergone significant changes to address student needs, the transmission of singing knowledge continues to follow the traditional master-apprentice model. Singing students typically seek guidance from experienced professionals, and it is widely acknowledged that the most effective way to master the art of singing remains through individualized instruction (O’Bryan & Harrison, 2014).

Many prominent authors have stated that singing is first an art, then a science, yet research on the subject has been predominantly scientific. However, according to Bunch Dayme (2009), “what is unique is that singing exists within an artistic realm; it is the emergence of physical functions from mental processes interconnected by aesthetic and subjectively expressive determinations.” The art of singing has both a technical/mechanical and an aesthetic dimension. Therefore, emphasizing both in vocal training is crucial for individuals to develop a strong foundation in both aspects and to improve their performance skills.

Rogers (2013) expresses the subject as follows: “The art of singing, in its entirety, is nothing more than the skill of using the human voice—the most wonderful of all musical instruments—correctly, beautifully, and effectively. When the act of singing is a passion and discipline in a person, the art of singing will be successfully elevated to its peak, both technically and aesthetically.”

Stark (2003) offers the following perspective: “There is an expressive power in the trained singing voice that I like to call ‘vocal aesthetics.’ This is the singer’s power to surprise, enchant, or impress the audience in a way that no other instrument can. Part of this expressive power, of course, stems from the presence of words that convey their own meaning. But beyond linguistic communication, there is another level of expression that arises from advanced vocal techniques. These techniques include timbre, interval and register characteristics, degrees of vocal intensity, vocal agility and richness, vocal onset, legato, portamento and other forms of articulation, trills, and rhythmic flexibility. Taken together, these techniques form what might be called a classic vocal ‘expression,’ as they are so well suited to the trained singing voice and distinguish it from both other forms of vocal use and instrumental expression.”

According to O’Bryan and Harrison (2014), the emergence of research in the fields of vocal physiology and acoustics revolutionized singing pedagogy, with the greatest change in understanding and teaching singing technique occurring in the last 30 years. However, while research in the field of vocal science can identify the biological and acoustic properties of the vocal mechanism, it cannot tell the individual how to conduct vocal training or how to teach singing skills. In this context, specific pedagogical strategies used by vocal trainers include the use of visual imagery through verbal instructions, the imitation of good vocal models, and gestural instructions. Singing techniques and pedagogy have changed in response to shifts in musical and societal demands. Observable changes in singing pedagogy from the beginning of the 19th century are clearly visible in the literature (O’Bryan and Harrison, 2014).

One of the most significant recent contributions to the literature on classical singing is Janice Chapman’s Singing and Teaching Singing (2006; revised 2011). In this work, Chapman explores a wide range of teaching and learning concepts, addressing not only the philosophy of vocal instruction but also the anatomical and acoustic foundations commonly found in vocal pedagogy texts. She further highlights the interdisciplinary nature of singing instruction. Chapman’s pedagogical approach integrates three principal perspectives, which she defines as “holistic, physiological, and incremental.” She asserts that effective teaching is always grounded in a guiding philosophy and maintains that “teaching is a creative process, never static, and must be continuously nourished by teacher/student interaction and development” (cited in O’Bryan & Harrison, 2014). In line with this, O’Bryan and Harrison (2014) summarize their educational philosophy as follows: “We base our knowledge on more than a century of scientific research into the mechanism of speech and value the age-old art of apprenticeship: learning by showing, doing, and transforming.”

According to McKinney (1994), singing is not a natural process but an art that requires mastery and highly developed muscle coordination. Therefore, effective training for developing singing skills necessitates the cultivation of the appropriate muscular structures through carefully planned exercise programs. The coordination of all the processes involved in singing is achieved through the harmony of a strengthened, alert, and lively mind, together with a healthy and vibrant body. Consequently, regular and systematically planned practice several times a week should be the goal for all singers. As Wilhelm Ehmann (1968) observes, “The singer uses his/her body both to sustain his/her life and to develop his/her art. He/she can never escape it, because his/her physical life either advances or hinders his/her artistic life” (cited in Buchanan, 2014).

Somatic educators study and teach the practical relationship between mind and body in action. The term somatic is derived from the Greek root soma, which is literally translated as “study of the body” (Johnson, 2009). In the field of education, Caplan (2009) explains that somatic refers to both the “study of the body in motion” and the “subjective experience of movement” (cited in Buchanan, 2014). In this context, Mingle (2018) emphasizes that the nature of singing is inherently somatic: mind, body, and spirit are embodied in the physical, expressive instrument that communicates through the body’s sound. Therefore, it is reasonable to conclude that principles underpinning somatic teaching methods can provide valuable tools for vocal instructors seeking to guide students toward realizing their intentions through efficient, coordinated, safe, and artistically expressive movement patterns.

A crucial factor in developing individuals' singing skills is the harmony between teacher and student. The impact of relationship on each individual's well-being is an important topic studied through social psychology theories. In this context, developing a healthy and successful relationship between teacher and student in one-on-one instrumental/vocal training can affect the student's performance outcome and benefit their learning processes (Kennell 2002; cited in Serra Dawa, 2014). Ideally, the relationship between teacher and student, which in most cases spans several years, should be a source not only of musical knowledge but also of inspiration, motivation, and psychological development. A healthy relationship provides space for the development and growth of all these qualities; however, in relationships lacking solid harmony, the student's development may be hampered. Therefore, care should be taken in pairing singing teachers and students (Serra-Dawa, 2014).

The vocal training process—traditionally centered on one-to-one instruction between instructor and student in the classroom—has long been, and should continue to be, a fundamental component of music education. Nevertheless, substantial pedagogical benefits may also be derived from the professional interactions that occur when vocal instructors participate in each other's practices when students share the experiences acquired through such exchanges (Harrison & O'Bryan, 2014).

Finally, to summarize as Harrison and O'Bryan (2014) also state: "While we may never know the origins of singing in all its complexity, we do know it to be part of the human condition. As folk singer Pete Seeger notes: "*Songs are funny things. They can slip across borders. Proliferate in prisons. Penetrate hard shells. I always believed that the right song at the right moment could change history* (1955). The singing teacher in the twenty-first century is therefore charged with an awesome responsibility. It is important for teachers and students to perform songs in a way that makes change possible through all kinds of songs. After all, that's what singing is."

In this context, the primary research question of the study is: "What is the importance of improving the singing skills of prospective music teachers, and what activities are carried out to achieve this goal?"

Sub-Problems

1. What is the importance of developing the singing skills of prospective music teachers?
2. What is the importance and role of acquiring technical knowledge and skills for vocal training at the theoretical and practical levels in developing the singing skills of prospective music teachers?
3. What is the importance and role of theoretical knowledge regarding the composer, period, genre, singing style, meaning of the lyrics if the work is in a foreign language, subject, story, etc., in developing the singing skills of prospective music teachers?
4. What is the importance and role of speaking exercises in developing the singing skills of prospective music teachers?
5. What is the importance of correct, beautiful, and effective interpretation of works in developing the singing skills of prospective music teachers, and what studies should be done for this purpose?
6. What is the importance and role of correct repertoire selection in developing the singing skills of prospective music teachers?
7. What is the importance and role of technological developments in developing the singing skills of prospective music teachers?

Purpose and importance of the research

The aim of the Vocal Training course, one of the most important and fundamental courses in the Music Education Undergraduate Program, is to develop the speaking and singing skills of prospective music teachers. Considering that the human voice is the most basic and natural instrument used in music education, and that singing skills are the aim and natural outcome of vocal training, providing prospective music teachers with sufficient and high-quality training in terms of knowledge and practice in the field of vocal training, and developing their singing skills, becomes even more important. Improving the singing skills of prospective music teachers will contribute to their being more equipped, confident, and successful both during their undergraduate education and in their future professional lives. Furthermore, while using their own voices correctly and setting a good example for their students, they will also teach and instill a love of singing in their students in the best way possible, while protecting their vocal health. On the other hand, music is the highest expression of art, and the highest expression of music is the art of singing. The fundamental element in art is emotion, and in the art of singing, there is a direct expression of emotions belonging to the human soul. Furthermore, it is a generally accepted understanding that music begins with singing and that singing is the most universal musical activity. Therefore, all these reasons make this study important. In this context, the aim of the research is to evaluate and make recommendations regarding the importance of developing the singing skills of prospective music teachers. To the extent that studies in our country are accessible, research in the field of voice training has generally focused on the fundamental processes of voice training, such as "posture, physical-mental and spiritual comfort, breathing, phonation, resonance, and articulation," as well as students' attitudes towards voice training lessons, motivation, self-efficacy levels, voice health, voice defects, voice training repertoire, voice exercises, imagery in voice training, approaches and methods used in voice training, examination of voice training curricula, etc. Therefore, since no studies on this subject could be found, it can be said that this study is the first of its kind. It is hoped that this study, which is considered important for improving the quality of the "Voice Training" courses in the Music Education Undergraduate Program, will be useful for voice trainers, students, and young academics, and will shed light on new research to be conducted.

Limitations

This study is limited to 2nd, 3rd, and 4th-year students enrolled in the Music Education Department of the Buca Faculty of Education at Dokuz Eylül University who have taken and are currently taking vocal training course. 1st-year students were not included in the study as they had recently started taking vocal training courses.

METHODOLOGY

Research Design

In this research, a literature review was conducted and a qualitative research method was used. A literature review is the reading, selection, and effective critical evaluation of existing information, thoughts, discussions, speculations, and findings related to the research area in order to achieve a specific goal (Ekiz, 2003). For a literature review, it is recommended to conduct a comprehensive examination, especially of the literature of recent years, to reveal the direction of global trends, and to access both domestic and international literature (Demirel, 1999). In line with these recommendations and the purpose of the present study, relevant national and international literature on the subject was examined.

"Qualitative research is research that follows a qualitative process aimed at presenting perceptions and events realistically and holistically in their natural environment, using qualitative data collection methods such as observation, interviews, and document analysis" (Yıldırım and Şimşek, 2004).

The research data was collected using the interview technique, one of the qualitative research methods. Briggs (1986) argues that the interview is the most common data collection method used in research in the social sciences. According to Briggs, the interview method is a highly effective method for obtaining information about individuals' experiences, attitudes, opinions, complaints, feelings, and beliefs (cited in Yıldırım and Şimşek, 2004). The interview, which is a technique for collecting data verbally, is a conversation organized for scientific purposes. A scientific interview needs to be carefully planned in advance (Arlı and Nazik, 2004).

The interview serves two main purposes: 1. To motivate the interviewee to provide complete and accurate answers, and 2. To eliminate biases stemming from factors such as social willingness or conformity. The interview is structured to create a social environment that facilitates free information exchange between two individuals (interviewer and interviewee) in order to reveal the interviewee's interests, views, attitudes, and behaviors. The interviewer must achieve these goals through interaction with the interviewee (Delmann, 1978; cited in Balcı, 2004). The success of the interview largely depends on selecting an unbiased sample, and this stage is a crucial task for the interviewer (Balcı, 2004).

This research also functions as a "case study." "A case study is an empirical research method used when examining a current phenomenon within its real-life context, where the boundaries between the phenomenon and its context are not clearly defined, and where multiple sources of evidence or data are available" (Yin, 1984).

In light of this information, the study defined the art of singing, singing skills and singing education by reviewing relevant domestic and foreign literature. Furthermore, based on the opinions of prospective music teachers, an attempt was made to identify the importance of developing singing skills and the studies that have been and should be conducted for this purpose.

Study Group

The study group of the research consists of 25 prospective music teachers studying in the Music Education Department of the Buca Faculty of Education at Dokuz Eylül University who have taken and are currently taking vocal training course.

Data Collection Process and Tools

In this study, a "semi-structured interview form for prospective music teachers" was used. The interviews were conducted with a total of 25 students studying in the Music Education Department of the Fine Arts Education Division at Buca Faculty of Education and taking vocal training courses. The semi-structured interview form prepared for data collection was evaluated by 4 vocal training experts, 2 music education experts, and 2 education science experts, and as a result of these evaluations, the number of items was determined to be 10. When preparing the interview questions, care was taken to ensure that they were clear and understandable, written in simple language, had a quality that could serve the purpose in terms of content, and that there was consistency between the questions. The data obtained by conducting interviews with each of the prospective music teachers were stored in the form of interview notes and audio recordings.

Data Analysis Techniques

The interview form, prepared based on expert opinions and piloted with 5 individuals, was administered to a total of 25 students. Interviews were recorded and transcribed using audio recordings and notes. The transcripts were analyzed using descriptive analysis methods. To ensure the validity of the analysis, the recordings were transcribed and analyzed by two separate researchers. To confirm the validity of the data, the data sources were revisited. In analyzing the interview data, the prepared questions were first categorized. Then, the responses to each question were divided into sub-dimensions based on frequency of use, and sample sentences related to each category were identified from among similar sentences.

FINDINGS

Findings Related to the First Sub-Problem

The first sub-problem of this research is as follows: What is the importance of developing the singing skills of prospective music teachers?

Category: The importance of singing skills

Question 1: In response to the question, "As prospective music teachers, what are your thoughts on the importance of developing your singing skills in terms of your education and future career?", all students emphasized that developing singing skills is crucial both during their undergraduate studies and for their future professional careers. The reasons they provided are as follows:

- Since the human voice is the only instrument all humans possess, all music teachers should be able to sing at a sufficient level.
- Because many subjects/fields in the education process require the use of the voice, knowing how to use it without damaging it is important and necessary.
- In the future, it will be easier to impart knowledge and skills to young students through singing.
- A music teacher will almost always need their voice and will always have to use it in their lessons. Furthermore, a teacher is a role model; therefore, a music teacher who uses their voice correctly and effectively will both be a good example for their students and motivate them.
- A music teacher should have a distinctive voice, whether singing children's songs to students in a family setting or in different social environments, or confidently singing in a choir they participate in as a music teacher.
- Possessing advanced singing skills equips prospective music teachers for their professional careers and plays a crucial role in fostering students' love for music.
- Singing, which is crucial for musical development, especially for children's musical development, is a skill that helps in conveying emotions, building self-confidence, and communicating.

Example sentences:

Interviewee 1: The human voice is the only instrument all humans possess. Therefore, all music educators need to be able to sing at a sufficient level. For example, when teaching world music to high school students, demonstrating it practically rather than just having them listen will provide much more lasting learning.

Interviewee 9: I believe that improving my singing skills is important both in the educational process and in my future professional life. Singing helps to convey emotions, build self-confidence, and communicate. During the educational process, it helps me both develop my own musical perception and gain mastery of different styles. In this way, I can introduce my students to different types of music in the future. I can also guide them in stage discipline and building self-confidence.

Interviewee 13: I believe that developing singing skills in music education programs is the most valuable and fundamental achievement a prospective music teacher should acquire. Singing training should be the priority in music education. Instrumental playing should come afterward, as many scientific studies argue that students who haven't acquired singing skills will also be unable to perform well on their instruments. During the educational process, students internalize much musical knowledge through singing, better understanding their own voice and learning what they can do. This process, on the one hand, improves musical reading and writing skills, supports inner ear development, and directly affects the student's ability to have a well-trained ear. Furthermore, developing singing skills helps prospective teachers use their voices more healthily, effectively, and clearly, contributing to their self-expression and building confidence in their performance. We can say that a music teacher candidate who understands the importance of singing will conduct their lessons with an approach that prioritizes singing throughout their professional life. This approach by the teacher will positively influence the attitudes of children and young people towards singing in music lessons in educational institutions. Therefore, prospective music teachers need to know and understand why acquiring singing skills is so important in this context.

Findings Related to the Second Sub-Problem

The second sub-problem of the research is as follows: What is the importance and role of acquiring technical knowledge and skills in vocal training at the theoretical and practical levels in developing the singing skills of prospective music teachers?

Category: Technical knowledge and skills

Question 2 a): Regarding the question, "What are your thoughts on the importance and role of theoretical achievements in vocal training in developing your singing skills? (human voice, its characteristics, voice types, voice formation processes "breathing, phonation, resonance, articulation", etc.), all students stated that these achievements are important. The reasons were expressed as follows:

- Music teacher candidates' awareness of their own voices and knowledge of vocal physiology will contribute to their more accurate performance of the art of singing.
- Knowledge of theoretical concepts related to vocal training will make individuals more conscious and contribute to the conscious and correct use of the voice.
- Conscious learning will accelerate and facilitate the learning process.
- It will ensure technical development.
- It will enable them to be more knowledgeable and aware of the human voice, its characteristics, anatomy, and physiology. Knowing these skills is especially important when teaching students as a music educator.
- It will enable them to become more competent and professional in the field of music teaching and vocal training.
- It will greatly contribute to the acquisition and development of singing skills in music teacher candidates.
- It will contribute not only to technical development but also to healthy and consistently correct vocal use.
- It will ensure the protection of vocal health and prevent damage to the voice.

Example sentences:

Interviewee 1: Instead of singing flatly and unconsciously, knowing the physiology of this will allow us to perform our art more accurately.

Interviewee 3: These theoretical gains are of great importance in terms of using the voice correctly and not damaging it. I think that breathing exercises and articulation, in particular, play a big role in terms of expressive power.

Interviewee 17: I believe that we need to know what we do well. If I want to improve my singing skills, I must first recognize the voice and understand how it is produced. Knowing the types of voices and their characteristics is very important in order to better understand my own voice. I think that just practicing is not enough to improve the voice; theoretical knowledge is very important to understand what we do, why we do it, its benefits, and to move ourselves forward.

Question 2 b): To the question, “What are your thoughts on the importance and role of acquiring technical knowledge and skills for vocal training at the practical level in developing your singing abilities? (Correct posture, mental, emotional and physical relaxation, ensuring correct breathing and breath control, establishing the correct voice-breath connection and producing the correct voice, achieving resonance, register transitions (head voice, mid-range and chest voice), articulation, etc.),” all students expressed the opinion that these acquisitions are important. Their reasoning was that it leads to the conscious, confident, healthy, relaxed, correct, beautiful and effective use of the voice. Furthermore, it was stated that this knowledge and these skills are absolutely essential for becoming better equipped as prospective teachers.

Example sentences:

Interviewee 7: I think learning technical skills through practice is absolutely essential for improving my singing skills. A natural and fluid voice is very important, but this naturalness is only possible with the right technique. For example, when I'm singing a song I love on stage, if I'm not in the right posture or I can't control my breath properly, my voice is immediately affected, which lowers my performance. Mental and physical relaxation are also part of it. Since we convey not only the voice but also the emotion, if I'm nervous or my body is tense, it immediately reflects in the song.

Interviewee 8: Actually, for me, technique is everything. We learn to sing comfortably with technique. Every sound we produce without technique damages our vocal cords. I think correct posture is very important, because slouching, in particular, reduces lung capacity, causing us to breathe less, which of course affects us greatly.

Breath support is everything in singing, in my opinion. If we don't breathe in the right place and don't use apoggio, our larynx stays high, which especially hinders us in high notes (the timbre is distorted or doesn't reach the high notes directly). The register is also a very important concept. Knowing and learning the correct transitions according to our own vocal timbre is crucial, because otherwise, we can't reach high notes without pushing the voice. Singing outwards, not clouding the voice internally, is also very important. For this, diaphragm support and breath control are essential.

Interviewee 17: I believe that acquiring technical knowledge and skills is the most important aspect of developing singing abilities. Correct posture, comfort, and proper breathing are fundamental and crucial, like the foundation of a building. Technical details such as resonance, articulation, and register transitions are skills related to singing quality. I think acquiring these skills is very important for vocal health and conscious singing.

Findings Regarding the Third Sub-Problem
The third sub-problem of the research is as follows: What is the importance and role of theoretical knowledge about the composer, period, genre, singing style, meaning of the lyrics (if the work is in a foreign language), subject (if any), story, etc., in developing the singing skills of prospective music teachers?

Category: Theoretical achievements
Question 3: In response to the question, "In vocal training classes, are students informed about theoretical topics such as the composer, period, genre, singing style, meaning of the lyrics (if the work is in a foreign language), subject (if any), story, etc.? What do you think is the importance and role of these knowledge acquisitions in developing your singing skills?", the vast majority of students stated that they were informed, one student stated that this information was not given in class, one student stated that it was given rarely, and one student stated that it was given occasionally when the opportunity arose. Their general thoughts on the importance and role of these achievements are as follows:

- Ensuring a more accurate mastery of the work being studied,
- Contributing to a correct interpretation by understanding, feeling, and reflecting the intended emotion of the work,
- Providing a confident, beautiful, and meaningful performance,
- Contributing to a better understanding of the work, the period, and what the composer wanted to convey. It also serves as an encouraging factor in learning about periods and composers,
- Ensuring an enjoyable performance by immersing oneself in the work,
- Contributing to musicality through a correct, beautiful, and effective performance.

Example sentences:
Interviewee 1: This information is generally overlooked in class. I think the conditions of the time the piece was written and the composer's intentions are factors that will make a difference during the performance.

Interviewee 8: I always thoroughly examine a piece before I start singing it. I even do it before listening to it. If the piece is in a foreign language, I think it's very important to know the meaning of the lyrics and, if applicable, the subject or story of the piece. It's briefly mentioned in our lessons, but I look into it in detail. Knowing these things makes a difference in my interpretation of the piece, because it's not just about producing the sound correctly, but also about conveying the emotion correctly. It would be very difficult to feel and reflect that emotion without knowing what the song is about. For example, knowing which period a piece belongs to also shapes our style of singing. There is a big difference, both technically and emotionally, between a Baroque piece and a Romantic piece. Similarly, knowing the composer, at least having an idea about their musical language, helps to interpret the piece more accurately.

Interviewee 14: During my student life, I rarely received information. But I consider it important and necessary. Because when you sing, you can direct your expressions in this way. Otherwise, you just act according to the feelings the melody evokes in you, and in that case, you only add your personal interpretation.

Interviewee 25: Information is provided. After learning this information, the works we sing gain more meaning.

Findings Regarding the Fourth Sub-Problem
The fourth sub-problem of the research is as follows: What is the importance and role of speaking exercises in developing the singing skills of prospective music teachers?

Category: Speech exercises in singing
Question 4: To the question, “What is the importance and role of speaking exercises in developing your singing skills? Do you do diction and articulation exercises in voice training classes, especially pronunciation and articulation exercises for songs written in foreign languages?”, all students stated that speaking exercises are important and the vast majority stated that they do these exercises. The answers given are generally as follows:

- Especially in works written in foreign languages, pronouncing words correctly and understanding their meaning directly reflects on the expression. Therefore, if the work is written in a foreign language, attention should be paid to pronunciation in order to perform it correctly.
- Being able to say words correctly ensures that the song is understandable and fluent.
- To prevent the work from losing its meaning, it is important to say it correctly without distorting its meaning and emotion.
- If a person's speech is fluent, their singing is also good.
- If the work is written in a different language, it is important to sing as well as native speakers to keep that culture alive.
- It ensures that the words are conveyed clearly and understandably to the listener during the performance, helping to accurately reflect the emotion and expression. These are important in strengthening the interpretation in the song.
- It contributes to a more technical way of singing.
- It is a useful preliminary exercise to focus on the work rather than pronunciation while singing.
- Through speech exercises, the difference between correct and incorrect pronunciation becomes very clear.
- Speech exercises are important for improving a person's singing experience, providing the listener with a more pleasant and aesthetically pleasing listening experience.
- These exercises contribute to the most accurate, beautiful, and effective performance of the piece.

Example sentences:
Interviewee 11: If our speech is fluent, our singing is also good. Before singing foreign songs, we always work on how to pronounce the lyrics. Even if the song is in a different language, it's important to sing it as well as the native speakers, because we are keeping that culture alive.

Interviewee 13: Yes, I definitely do articulation exercises. I find "declamation opera" particularly helpful. That is, in the sense of "oratory" or "expression," it refers to works where the words are emphatic and impactful. Trying and practicing this helps strengthen the chest and middle registers in particular. In singing, speech exercises play a role in seeing the text as a whole and understanding the words as a whole. It ensures that the words are conveyed clearly and understandably, and helps the performance accurately reflect emotion and expression. These exercises are important in strengthening song interpretation.

Interviewee 17: Speech exercises are very helpful for using the voice in the correct position and with correct pronunciation, and we can clearly distinguish between right and wrong. In foreign language works, I think it's very important to do speaking exercises before performing a song to ensure correct pronunciation. It's a useful preparatory exercise to focus on the piece rather than the pronunciation while singing.

Interviewee 19: I frequently did articulation and diction exercises with all the teachers I worked with. It's not just about singing; it's something we need to pay attention to in our daily lives when communicating with people, and also when we serve as role models for our students as teachers.

Findings Regarding the Fifth Sub-Problem
The fifth sub-problem of the research is as follows: What is the importance of correct, beautiful, and effective interpretation of musical works in developing the singing skills of prospective music teachers, and what studies should be conducted for this purpose?

Category: The importance of interpretation and interpretation studies
Question 5: The answers given to the question, "What are your thoughts on the importance of correct, beautiful, and effective interpretation of musical works and studies aimed at this purpose in developing your singing skills?" are generally as follows:

- To show the necessary respect to both the composer and the audience, every task should be performed with due diligence. Therefore, a good interpretation of the works is crucial.

- It helps the individual to gradually develop and solidify their own ideas and feelings.
- Singing is not just about playing the correct notes and singing the song correctly, but also about conveying emotion.
- A good interpretation highlights the beauty of the work, giving the listener more pleasure.
- Interpretation plays a huge role in reflecting the spirit and meaning of the work.
- One of the most important aspects of musical expression is understanding the message of the work, accurately reflecting its spirit and emotion, interpreting the work well, and conveying this to the listener. Therefore, the better a work is interpreted, the more beautiful an impact it leaves.
- When performing a work, technical equipment must be integrated with emotional and artistic expressive power.

Example sentences:
Interviewee 1: Beyond singing, I think every job should be done with dedication so that we can show respect to both the composer and the audience.

Interviewee 9: Simply singing a piece correctly technically is not enough to fully bring it to life. One of the most important aspects of musical expression is being able to correctly interpret the spirit and emotion of the work and convey it to the listener.

Interviewee 13: I think that improving singing skills is more than just singing correctly and with the right technique. For the correct, beautiful, and effective interpretation of works, it is necessary to understand the message that the composer and lyricist want to convey, to adhere to the notation, and to pay attention to the rhythm and tempo. Clear and understandable expression of the lyrics is important for the emotional comprehension of the work. Each song genre has its own unique interpretation style. For example, the way an opera aria is interpreted differs from the way a folk song is interpreted. Furthermore, for a singer to understand their own emotional world, develop their expressive abilities, and use their body as an instrument, it is beneficial to practice physical relaxation and mindfulness exercises.

Interviewee 15: …When performing a piece, technical equipment must be integrated with emotional and artistic expressive power.

Question 6: "In vocal training classes, which of the following exercises do you use to improve the interpretation of musical works, and what are your thoughts on their effectiveness? What are your brief opinions on each item? If you have any examples from your own life, would you share them?"

- Expressing musical phrases through speaking exercises
- Nuance exercises for better expression of musical phrases
- Knowing/learning the meaning of song lyrics written in a foreign language
- Gaining knowledge about the subject/story of the works
- Expressing the works with musical expression in accordance with the lyrics
- Establishing a connection between the individual's life and memories and the emotion created by the lyrics and melody of the works
- Imagination exercises related to the emotion created by the lyrics and melody of the works
- Examining the works from historical, social, and philosophical perspectives

In response to this question, 17 students generally stated that they did the first 6 exercises, 6 students stated that they did all but the last item, and 2 students stated that they did all the exercises.

Example sentences:
Interviewee 10: Expressing musical phrases through speech exercises: We do this in class. This exercise is effective in making the musical phrase more meaningful by standardizing it in the mind and performing it more appropriately. Nuance exercises for better expression of musical phrases: We do this in class. When we break the piece down into parts and establish the nuances sentence by sentence, the overall piece is expressed more quickly and beautifully. Knowing/learning the meaning of song lyrics written in a foreign language: We do this in class. Knowing the meaning of the lyrics allows us to better express the emotions conveyed by the lyrics when performing the piece. For example, if a piece contains sentences about longing for a loved one, when I read the piece, I amplify the feeling of longing for my loved one and reflect this emotion in my voice and manner of singing. Gaining knowledge about the subject/story of the pieces: We do this in class. Knowing what the piece is about helps us know what emotion to sing it with. When the subject/meaning is known, singing as if you are living

the story helps to add a much more beautiful interpretation to the piece. Expressing the work through musical storytelling in accordance with its lyrics: We do this in class. It's important for conveying emotion to the listener. Connecting the individual's life experiences and memories with the emotion created by the lyrics and melody of the work: I experienced this in our last lesson. While working on the piece "Nina," with my teacher's guidance, knowing where and how to make nuances, and with the emotion that comes from knowing the story of the piece, I interpreted the work in its most beautiful form, and my eyes truly filled with tears while singing it.

Interviewee 13: I do all of these exercises. In speech exercises, I find working through passages particularly helpful. For example, there's a word in the piece that reflects a certain emotion. I always learn the meaning of that word (if it's in a foreign language), and the melody and lyrics are usually composed/written to match that meaning. In one passage, the word "piango" means "I cry," and the expression there requires the necessary emotion and expression in the voice. For this, understanding the lyrics written in a foreign language is very important, and the musical expression must be appropriate. Regarding expressing musical phrases well, especially if the piece is in the style required by the Bel Canto period (e.g., Bellini, Donizetti, Verdi), I focus especially on being able to perform the messa di voce in a distinct and beautiful tone. I think about these nuances and perform them within the piece. Learning the subject matter of the piece (especially if it's an opera aria) is important for accurately portraying the role. If I'm singing a mezzo-soprano role, knowing the character and attitude of that role actually represents an understanding that goes beyond just vocal technique. Because details like whether the mezzo-soprano is singing in a lyrical or dramatic style in the piece allow me to perform the work well and accurately. I think I do more imagery work in Tchaikovsky's romances than in the lyrics. For example, I visualize a place, a smell, colors, etc., mentioned in the song. A good example of this is Schubert's Erlkönig, with a poem by Goethe. This song has an effect that directly brings out imagery. I also value the historical, social, and philosophical examination of the work. Because during the period in which the works were composed, composers were influenced by and interacted with intellectual people in other fields of art (poets, painters, etc.). Therefore, learning what the composer wanted to convey, who they communicated with at that time, and who they read allows us to connect with the work more deeply.

Question 7: The answers to the question, “In vocal training classes, what kind of exercises do you do yourself to improve your interpretation of musical pieces? What are your thoughts on the impact and benefits of these exercises in developing your singing skills? (Participating in scientific activities, reading publications on the human voice and its training, listening to different music and singing styles, attending concerts, etc.). If you have any examples from your own life, would you share them?” are generally as follows:

- Performing technical exercises focusing on breath and voice,
- Trying to understand the work through its lyrics and musical structure,
- Conducting research on the work and the composer,
- Listening to the work performed by many different artists,
- Listening to music and songs in different styles,
- Attending musical events, operas, and concerts,
- Reading articles about the human voice, participating in scientific events, and attending masterclasses,
- Trying to learn how artists use techniques by listening, watching, and observing. • Actively following many opera pages on social media, communicating with opera artists, and learning about their techniques.

Example sentences:
Interviewee 8: Honestly, I've done a lot in this area. I'm someone who closely follows the events of the Izmir State Opera and Ballet and the Ahmet Adnan Saygun Art Center, and I frequently attend concerts (especially opera performances, opera nights featuring arias by specific composers, symphony orchestra concerts, etc.). Attending these concerts and listening attentively has definitely allowed me to better understand certain things in my field, sometimes seeing what shouldn't be done, and sometimes learning things that can improve my technique. I've also been watching the semi-finals and finals of some vocal competitions open to the public since I started working in this field, and this has also helped me better understand what I should or shouldn't do in my area. I also actively follow and watch many opera pages on social media. This again allows me to find things that can help me improve myself. Sometimes I especially contact foreign opera singers and ask them questions about their techniques. I've actually incorporated this into my life, so I'm very open to anything that can help me improve myself and I strive to learn as much as I can.

Interviewee 13: I listen to numerous examples of the works I'll be studying on digital platforms, and this greatly helps me fully learn a song. This way, I see different interpretations and learn about alternative interpretations I might prefer. I read scholarly articles to learn about the historical background of the art of singing. This allows

me to understand in detail the history and evolution of singing, which holds such an important place in human life. In addition, I attend masterclasses and opera performances to improve my knowledge and skills in singing.

Interviewee 24: To perform the piece better, I listen to different interpretations, research the work and its composer. I believe that because I understand the emotion and the piece better, I can perform it better.

Findings Related to the Sixth Sub-Problem
The sixth sub-problem of the research is as follows: What is the importance and role of choosing the right repertoire in developing the singing skills of prospective music teachers?

Category: Repertoire in singing
Question 8: The answers given to the question, "What is the importance and role of choosing the right repertoire in developing your singing skills? (When you evaluate the selection of works in the right key for the appropriate repertoire, the selection of works suitable for your age and maturity level, the ease/difficulty level of the works in terms of melody, rhythm and harmonics, etc.). If you have had a positive or negative experience in this regard, would you share it?" are generally as follows:

- In repertoire selection, readiness, vocal range, vocal timbre, age, experience, musical skills, and technical level are important. Choosing the right repertoire will allow the student to both recognize their own vocal characteristics and come to class more willingly.
- Choosing the right repertoire is crucial for correctly interpreting and doing justice to a piece from a technical and musical perspective. Incorrect choices can cause the best aspects of the piece to be lost.
- To interpret a piece beautifully, one must first embrace and love it. This allows for a more accurate and effective performance.
- Choosing the wrong piece can discourage a student from continuing the lesson and seriously damage their voice, resulting in a performance below average.
- Choosing a piece that suits one's taste and talent will allow them to interpret the piece comfortably and confidently, and with enjoyment and willingness.
- Pieces that allow a person to showcase their talent are important for both the singer and the listener.
- Maintaining vocal health is important for using the voice more accurately, comfortably, and effectively.

Example sentences:
Interviewee 1: I think that different types of songs should be included in vocal lessons so that the student is more motivated to come to class and can more easily find their own vocal timbre. My teacher took an innovative approach in this regard, and because of this, I deliberately left the class and took vocal lessons for another half-semester.

Interviewee 13: Choosing the right repertoire is actually more important in terms of not harming the student's voice. Determining the area where the student feels most comfortable with their voice and can perform at their best is the most important step in developing singing skills. Choosing pieces that exceed the student's readiness level and push their vocal limits can both discourage the student from the lesson and seriously damage their voice. Also, giving a melodically and harmonically difficult piece to a student who is not good at sight-reading and solfège can cause the student to lose interest and feel inadequate. This is completely outside the goals of vocal training lessons.

Interviewee 14: I think choosing the right repertoire is extremely important. A repertoire can enhance or diminish the value of a voice, or even harm it. I enjoy singing when the piece suits my voice. Singing an unsuitable piece feels like an obligation and is very boring. Also, my voice gets hoarse and my desire to sing decreases.

Interviewee 23: Choosing the right piece is important both for development and motivation, and for vocal health.

Findings Regarding the Seventh Sub-Problem
The seventh sub-problem of the research is as follows: What is the importance and role of technological advancements in improving the singing skills of prospective music teachers?

Category: Technology in singing
Question 9 a): The answers given to the question, "What is the importance and role of technological advancements in improving your singing skills?" are as follows:

- With technological advancements, access to many works, working techniques, and theoretical knowledge has become easier through quick and efficient methods.
- It provides the opportunity to instantly listen to a song of interest and easily learn all kinds of information about it.
- It provides the opportunity to easily learn all kinds of information related to vocal training.
- It contributes to accessing the sheet music of works.
- It allows one to go beyond the sheet music and see different performance styles, listen to examples of works from many master artists, and develop critical skills.
- Since many songs also have karaoke versions, it allows one to sing along.
- Digital platforms contain all the necessary information and unlimited content to improve vocal training skills.
- Thanks to online lessons, students can improve their voice and singing skills with an instructor through online classes.
- Voice recording devices, computer programs, and various applications allow for recording, listening, analyzing, and identifying errors, as well as enabling both analysis and evaluation by improving the voice.
- Digital applications allow for regular practice sessions.
- Applications enable vocal exercises and allow for working on rhythm and intonation by singing with digital platforms.
- Technology facilitates access to opera, concerts, recitals, etc., today.
- It offers access to operas not staged in our country, as well as providing access to a wealth of useful information related to opera and vocal training. It also allows for following and learning from people involved in this field.
- It allows for practice with accompaniment at the desired key and time, without the need for an accompanist, thanks to applications that can change the key.
- If a work is in a foreign language, it provides immediate access to its translation.
- Students can easily learn the correct pronunciation and interpretation of the work they will be studying through video recordings, and improve their own interpretation in this area.

Example sentences:
Interviewee 5: One of the most beneficial effects of technological advancements on singing skills is the ability to listen to examples of musical works online. We can listen to examples from many master artists. It also contributes to accessing sheet music. Whether it's the stories behind the works or their staging, accessing them is much easier today thanks to technology.

Interviewee 7: Recording my voice and listening to myself helps me identify my mistakes. I can do vocal exercises with apps, and I can work on rhythm and intonation by singing with digital backing tracks.

Interviewee 12: I choose a piece, and without needing an accompanist, I can open it on YouTube at any time I want and practice in any key I want. If the key I want isn't on YouTube, there are apps that can change the key. If a piece is in a foreign language, we can immediately access its translation.

Interviewee 14: There is no area that technology hasn't entered or affected. We encounter technology all the time. It can be harmful or beneficial. It depends on how you use it. When used for educational purposes, it can be extremely beneficial. Voice recording programs, instructional videos, access to distance learning, online seminars, and recently even AI-based programs that provide feedback... I believe that nowadays, only those who are intrinsically unwilling to learn will fail to do so.

Question 9 b): When evaluating the responses to the question, “What kind of activities do you do with your instructor or on your own in class regarding the use of technology? What are your thoughts on the usefulness and necessity of these activities? Can you share an example from your own experience?”, the common opinion reached by prospective music teachers is this: Technological applications are very important for the development and diversification of students' singing skills because they offer a very useful, fast, and practical way.

The responses to the studies conducted are as follows:

- Access to sheet music and hard-to-find notes.
- Ability to select musical works.
- Ability to record and listen to audio recordings to identify and correct errors.
- Using websites to learn more detailed information about the works studied.

- Researching the history, composer, subject, story, and Turkish meaning of the work if it is in a foreign language.
- Studying works by listening to different performers to improve technical and interpretive skills.
- Developing critical thinking skills by listening to many artists.
- Watching videos on vocal training techniques and incorporating and re-applying them in the exercises done in the lessons.
- Combining the work done in the lesson with video resources found on the internet related to the work.
- Doing metronome exercises, breathing and vocal exercises from online resources.
- Watching excerpts from operas, concerts or solo performances with the instructor or on their own in the lessons.
- Correcting the pronunciation of songs written in foreign languages by listening.
- Performing the piece with piano accompaniment videos.

Example sentences:
Interviewee 3: In classes, my instructor and I analyze the piece I'm working on by listening to different performers. We also use technology by watching artists' performances to better understand the correct techniques. In my own work, I can incorporate techniques, exercises, and breathing exercises I see on the internet and through digital applications into my own practice.

Interviewee 4: I record my voice and listen to it, and I do exercises from online resources. Metronome and breathing exercises are also very helpful.

Interviewee 11: We watch concerts or solo performances. If there's an unknown song I want to sing, I can have them listen to it. If there's a part of the pronunciation of foreign songs that we're unsure about, we can listen to it. We can find hard-to-find notes.

Interviewee 18: My instructor and I use technology to understand the meanings of the lyrics of foreign works and to select pieces. When I do individual work, I use videos related to vocal exercises and techniques. I believe it's right to keep up with advancing technology and to use its benefits in my studies.

Interviewee 20: My instructor and I focus quite a bit on correct storytelling techniques and interpretation. I also individually watch video resources related to the work that I can find online, and I combine that with what we do in class.

Category: Opinions and suggestions
Question 10: Regarding the question, "Do you have any different opinions and suggestions on how to improve your singing skills?", 19 students provided opinions and suggestions, while 6 students did not express any opinions or suggestions. The opinions and suggestions generally included the following:

- Since the process of voice production and singing are abstract concepts, they are very difficult to perceive and understand. Therefore, variety and imagination are very important in vocal training. Thus, the method may involve combining the experiences and techniques of many individuals and/or instructors.
- First and foremost, correctly perceiving and internalizing musical pieces is the most important step. Technical knowledge and exercises can be applied correctly on this foundation.
- Learning different and varied exercises for vocal training will be beneficial.
- Body and breath awareness exercises can be incorporated into vocal training.
- Information about vocal health and protection can be provided.
- The organs used in singing should be taught not only through explanation but also through demonstration (posters, skeletons, etc.).
- Digital resources related to vocal training can be recommended to students.
- A supportive and non-judgmental learning environment, along with supportive and encouraging approaches from teachers, will directly impact the development of self-confidence.
- Stage experience is also very important in the vocal training process. Therefore, giving students more opportunities to sing in front of an audience will significantly improve their self-confidence and interpretation.
- For students with stage fright, psychological support, breathing exercises, or group-based trust-building activities can be helpful.
- Singing is not just about technically correct singing; it also expresses creative and emotional meaning. This should be considered in vocal training classes.

- After the foundations of vocal training are properly laid and a sufficient classical repertoire is established, if appropriate, genres that the student enjoys listening to and singing can be included. In repertoire selection, if different pieces are chosen for each student, within their comfortable vocal range, suitable to their vocal characteristics, allowing them to showcase their voice, and which they can sing with enjoyment and acceptance, and where the pleasure of singing is prioritized, their enthusiasm and interest in singing can be increased.
- As with piano lessons, vocal training classes should include one Baroque, one Classical, one Romantic, and one Turkish piece each term. This ensures a thorough understanding and differentiation of the period of the works.
- Listening to and learning about songs from different cultures is also very important. This can enrich musical expression.
- Awareness of one's own voice is crucial, therefore regular and disciplined practice is essential for voice development. Alongside this practice, patience and energy can be important for improvement. Anyone who wants to improve themselves needs to research, listen, and practice extensively.
- Regardless of age and experience, learning and continuous development are ongoing. Therefore, one should always strive to add more to their knowledge and gain new experiences.
- Considering that the voice training process is long-term, lessons should not be crammed into a short period but should continue uninterrupted throughout the training years.

Example sentences:
Interviewee 6: To improve our singing skills, listening to songs from different cultures and getting to know those cultures is very important. This enriches our musical expression.

Interviewee 7: I think stage experience is just as important as technical training for improving singing skills. Being given more opportunities to sing in front of an audience significantly improves self-confidence and interpretation.

Interviewee 14: I think variety and imagination are very important. Since the process of voice production and singing are abstract things, it is a very difficult subject to perceive and understand. It often happens through imagination. Therefore, I don't think there is a single technique. The most suitable technique is the one that suits you best, that doesn't harm your voice, and that helps you produce the desired sound. The method that helps me the most is listening to the experiences and techniques of many people and/or instructors. With their perspectives, I create a broader picture and feel a sense of mastery.

Interviewee 17: For some students, the development of singing skills is not limited to technical training; it also requires strengthening self-confidence. Choosing the right repertoire, understanding the characteristics of the voice, and regular practice, along with a supportive and non-judgmental learning environment, are crucial. Teachers' supportive and encouraging approaches directly impact the development of self-confidence. Furthermore, considering that this is a long-term process, I believe that lessons should not be crammed into a short period but should continue uninterrupted throughout the academic years.

CONCLUSION, DISCUSSION AND RECOMMENDATIONS

The aim of this research is to evaluate and make recommendations regarding the importance of developing the singing skills of prospective music teachers. Based on interviews with prospective music teachers, the following conclusions were drawn from the research findings:

- Developing the singing skills of prospective music teachers is crucial both during their undergraduate education and in their future professional lives. Because the human voice is the only instrument all humans possess, all music teachers should be able to sing at a sufficient level.
  In developing the singing skills of prospective music teachers:
- Acquiring technical knowledge and skills related to vocal training at both theoretical and practical levels is important and necessary.
- Gaining knowledge of theoretical topics such as the composer, period, genre, singing style, meaning of lyrics (if the work is in a foreign language), subject matter, and story of the work is important.
- Speaking exercises are important and necessary.
- Interpretation exercises aimed at correct, beautiful and effective performance of works are important.
- Choosing the right repertoire plays a significant role.
- Technological advancements play an important and influential role.

When the studies mentioned below regarding interpretation are evaluated, it is concluded that the vast majority of students generally completed the first 6 studies, 6 students completed all but the last item, and only 2 students completed all the studies.

- Expressing musical sentences through speaking exercises
- Nuance studies for better expression of musical sentences
- Knowing/learning the meaning of song lyrics written in a foreign language
- Gaining information about the subject/story of the works
- Expressing the works with a musical narrative in accordance with the lyrics
- Establishing a connection between the individual's life and memories and the emotion created by the lyrics and melody of the works
- Imagination studies related to the emotion created by the lyrics and melody of the works
- Examining the works from historical, social and philosophical perspectives

The activities undertaken by prospective music teachers to improve their interpretation of musical works include:

- Practicing breath and vocal techniques
- Trying to understand the work through its lyrics and musical structure
- Conducting research on the work and its composer
- Listening to the work performed by many different artists
- Listening to music and songs in different styles
- Attending musical events, operas, and concerts
- Reading articles about the human voice, participating in scientific events, and attending masterclasses
- Trying to learn how artists use techniques by listening, watching, and observing
- Actively following many opera pages on social media, communicating with opera artists, and learning about their techniques

Discussion

The first sub-problem of the research, "What is the importance of developing the singing skills of prospective music teachers?", indicates that developing the singing skills of prospective music teachers is important both during their undergraduate education and in their future professional lives. Because the human voice is the only instrument all humans possess, all music teachers should be able to sing at a sufficient level. Singing is crucial for musical development, especially for children. It is also a skill that helps in conveying emotions, building self-confidence, and communicating. In this context, possessing advanced singing skills will equip prospective music teachers for their future careers and play a significant role in instilling a love of music in their students. Furthermore, since a teacher is a role model, a music teacher who uses their voice correctly and effectively will both set a good example for their students and motivate them.

According to Rao (1988), the art of singing requires mastery. Mastery in the art of singing is a way of “hearing the music, imagining the music, knowing the music, and performing the music.” In this context, in order to produce music with the human voice, all parts of the vocal system must function as an artistic whole. The concept of the art of singing as a form of vocal art should be broad enough to encompass every aspect of vocal production, from acquired habit through training to practice-based skill, practice-based technique, and artistic creation. All of these are forms of learnable procedural knowledge, that is, ways of knowing how to do something. Art is tied to the practice of skills, no longer seen as isolated skills, but as technical skills. The art of singing as a vocal art can be defined as a fusion of action and awareness; in this sense, it is a unity of skill and understanding.

The most important means of communication and natural instrument of human is one's own voice. Using and preserving this voice through proper habits is possible with vocal training that begins at a young age. The fundamental goal of music education is to train a child's voice and enable them to sing correctly. In line with the goals of music education, the teacher's own voice is the most frequently used tool in the process of creating behavioral changes in the student. One of the basic principles of the Kodaly method, a contemporary approach to music education, is the idea that everyone can use their own voice as the most suitable, accessible, and inexpensive musical instrument in learning and developing music. According to Kodaly, the human voice is the foundation of music. Singing prepares the ground for and helps in the development of other musical instruments. Not all children have the ability to play an instrument, and it is difficult for everyone to own one. However, every healthy child possesses their own voice, their most beautiful and natural instrument (Yiğit, 2000). In this context, it is extremely important for a music teacher, who serves as a role model for their students, to be a competent vocal trainer who sings well.

Regarding the second sub-problem of the research, "What is the importance and role of acquiring technical knowledge and skills in vocal training at the theoretical and practical levels in developing the singing skills of prospective music teachers?", the common opinion of prospective music teachers is that these acquisitions are important. These acquisitions will make prospective music teachers more conscious and equipped, helping them to recognize their own voices, understand vocal physiology, and ultimately perform the art of singing more accurately. Conscious learning will accelerate and facilitate the learning process by ensuring technical development. It will also contribute to the healthier and longer-lasting use of the voice. In conclusion, it will greatly contribute to the acquisition and development of singing skills, thus enabling individuals to become more competent and professional in the field of music teaching and vocal training.

According to the scientific approach, the act of singing, which is a "coordinative process," involves both acquiring knowledge about the voice and practicing by adapting the voice as needed. Similarly, Appleman (1986: 9), who defines singing as "disciplined expression," states: "Psychophysically, artistic singing is a dynamic act of instantly coordinating the physical sensations of respiration, phonation, resonance, and articulation to transform these processes into disciplined expression."

According to Howard (1982), the ability to sing artistically requires a combination of scientific understanding of the voice and practical ability to control it. However, while knowledge of the lungs, vocal cords, etc., can contribute to successful tone production, the application of this knowledge should be realized through vocal practice rather than verbal explanations. This is the difference between singing theory and the art of singing. Singing theory requires only scientific knowledge. The art of singing, on the other hand, requires the ability to put scientific knowledge into action. Such sensitivities are achieved through instantaneous awareness of vocal processes, including:

- awareness of vocal functions (breathing, phonation, resonance, and articulation);
- assessment of changes occurring during singing (changes in pitch and rhythm depending on changes in vocal cord movement); and
- the ability to replicate these changes.

According to Sabar (2008), a person’s vocal potential and talent can be developed through high-quality vocal training with a skilled instructor, a solid theoretical foundation, and mostly correct and appropriate practice. In this context, it is advisable to follow a progression from easy to difficult and from simple to complex, providing explanations and exercises that the trainee can comprehend. Given that the human voice possesses a lively, natural, and sensitive structure, its training requires equally sensitive, careful, dedicated, and patient work. In this regard, Cynthia Hoffmann (2003) emphasizes that the role of a vocal coach is to “help young students become artists who can express themselves more freely in terms of musical performance” and to support their musical performance and expression skills physically, spiritually and mentally” (cited in Jung, 2010). Furthermore, Santelli (2023) argues that the technical objectives set by instructors in voice training should ideally encompass broad realities related to vocal function and, ultimately, avoid imposing personal ideals that could limit individuality.

The third sub-problem of the research, "What is the importance and role of theoretical knowledge regarding the composer, period, genre, singing style, meaning of lyrics (if the work is in a foreign language), subject matter, story, etc., in developing the singing skills of prospective music teachers?", indicates that these acquisitions are important. These acquisitions contribute to a more accurate understanding of the work, allowing for a more precise interpretation by understanding, feeling, and reflecting the intended emotion. Therefore, it leads to a confident, beautiful, and meaningful performance. Besides contributing to a better understanding of the work, its period, and the composer's intentions, it also encourages learning about different periods and composers. Furthermore, it contributes to an enjoyable performance by immersing oneself in the work. In short, it contributes to musicality through a correct, beautiful, and effective performance.

According to Hemsley (1998), when working on a song lyric for a good interpretation, the following should be considered: “meaning, basic rhythm and beats, the sounds that make up the lyrics, and where new musical phrases emerge (each musical phrase is a new musical expression).” Meaning is especially important for songs written in foreign languages. Therefore, songs written in foreign languages need to be translated accurately and carefully. The pronunciation of every word in the song should be understandable, and its meaning should be pleasing through aesthetic expression. At this point, it is important to answer the following basic questions:

- What is the poet or lyricist saying? What is the content of the text and what emotions are expressed? Is the singer telling a story or acting out a scene? Or is he/she expressing personal thoughts? Is the song a

prayer or a meditation? The singer should categorize the song and have a clear mind about the general nature of the message he/she wants to convey to the listener.

- Who is singing the song? Is it a clearly defined character, as in an opera or musical, or just the singer? At this point, it is important to remember that most songs express the feelings and thoughts of young people.

- To whom is the song being sung? This is sometimes clearly evident from the lyrics, as in opera, but not always. Since singing is a form of communication, there must be not only someone communicating, but also someone being communicated with. This could be the world in general, God, a friend, a lover near or far, a hated enemy, a direct audience, or the singer themselves. Ultimately, the listener must be involved.

- What is the mood and atmosphere in the song? The range of moods is almost infinite. If singers have trained themselves to convey different moods accurately and clearly with their voices, they can allow themselves to be guided by intuition and imagination. However, they must know the mood they are expressing. If they have only trained themselves in vocal technique, then their only tools are variations in musical nuances and lyrics.

According to Webb (1946), every song has a message it wants to convey, and whether expressed through music or words, recognizing this message is necessary for faithful interpretation. Once this recognition is achieved, the outline of one's interpretation emerges. Then the song must be placed in the correct category. To what period or style does it belong? Should it be considered an atmospheric or melodic song? Is it narrative, descriptive, or subjective in mood? Does it have any extraordinary musical features? What is its structure? Where are its most important points? Where is the climax, and what part of all this does the accompaniment encompass? One must ask oneself all these questions, returning to the original repeatedly. Furthermore, it is important to find answers to the questions, "What does this song want to convey, and how can I express it, how can I use my voice, how can I look, how can I feel to reveal the meaning of the song?" At this point, artistry and technique must work together.

Webb (1946) states that the concept of style, which is difficult to understand, largely depends on the ability to place a song in the correct category by evaluating its origins and historical background. Examining the development of a song and understanding how the composer and poet interacted with each other in different periods is indispensable. In this context, informing the student about the selected works (the composer, genre, period, style of performance, culture to which it belongs, subject matter, meaning of the lyrics, musical structure, etc.) in vocal training classes is important in terms of laying the foundation for a much more effective training process focused on trust, harmony, and success between the vocal coach and the student (Ekici, 2020).

The fourth sub-problem of the research, "What is the importance and role of speaking exercises in developing the singing skills of prospective music teachers?", indicates that speaking exercises are important and necessary in improving singing skills. Speaking exercises contribute to making the song understandable and fluent. Especially in works written in foreign languages, correctly pronouncing words and knowing their meaning will directly reflect in the expression. To preserve the meaning of the work, it is important to sing correctly without distorting its meaning and emotion, especially if the work is in a different language; singing as well as native speakers is crucial for keeping that culture alive. Speaking exercises are also important for enabling a person to offer the listener a more pleasant and aesthetically pleasing listening experience, both technically and emotionally, while singing.

A music educator is someone who instills, develops, and serves as a role model for musical behavior. One of the most important tools for developing musical behavior is singing. Singing correctly, beautifully, and effectively depends on speaking correctly, beautifully, and effectively. From this perspective, a good music educator is expected to have a good command of their native language, a good understanding of the phonetics of Turkish and foreign songs, and to exhibit language-enhancing behaviors while teaching students songs. In this context, the education provided in institutions that train music teachers should address speaking and singing together, and the necessary skills should be imparted to individuals at both theoretical and practical levels. The key to singing is ensuring that the lyrics reach the audience clearly and understandably. Achieving this will balance the perfection of diction with performance skills. Therefore, a good singer should be able to use their voice both as a musical instrument and as a means of communication.

Although there are differences between speaking and singing, the basic mechanism, especially the physical processes, are the same, so singing is, in a sense, called extended and enhanced speech. Individuals skilled in both

can easily and comfortably transition from one to the other. Considering speech as the foundation of singing, many of the basic principles of singing, especially those related to posture (correct and balanced stance in accordance with the body's anatomical structure), breathing, and articulation, can be learned during speech at a comfortable vocal level. On the other hand, the desire and enjoyment of singing are a great aid in developing and maintaining rich, powerful, and beautiful speech sounds (McKinney, 1994).

According to Vennard (1967: 184), the word diction, which means choosing the right words to express thought, is used in singing to describe the best articulation of words. In this context, the subject can be approached from the perspective of clear and understandable pronunciation and good expression. To ensure correct pronunciation in a song, sounds should be produced as close to natural speech as possible, and care should be taken to ensure that each syllable is clear and audible. This requires being free from the anxiety of legato (singing or playing the notes of a musical piece without interruption), mental and physical relaxation, and controlled use of breath.

Kagen (1960) approaches the subject as follows: In order to achieve correct, beautiful, and effective expression in speaking and singing, it is very important for individuals receiving vocal training to use language correctly and to pronounce words clearly and intelligibly. Therefore, the fundamentals of speaking should definitely be considered a priority subject in the music education undergraduate program and should continue from the first year until the completion of the education.

The fifth sub-problem of the research, "What is the importance of correct, beautiful and effective interpretation of musical works in developing the singing skills of prospective music teachers, and what studies should be conducted for this purpose?", indicates that interpretation studies are important in developing singing skills. A good interpretation is important in showing due respect to both the composer and the audience. Singing is not only about correctly reading the notes and the song, but also about reflecting emotion. In this context, a good interpretation highlights the beauty of the work, leading to greater enjoyment for the listener. One of the most important aspects of musical expression is understanding the message of the work, correctly interpreting its spirit and emotion, and conveying this to the listener. Therefore, the more beautifully a work is interpreted, the more beautiful the impact it leaves. In short, when performing a work, technical skills must be integrated with emotional and artistic expressive power.

In general, technique, interpretation and expression are a whole in the performance of musical works. Therefore, in developing the singing skills of prospective music teachers, it is important and necessary for them to acquire interpretive and expressive skills in addition to technical knowledge and skills in vocal training at the theoretical and practical levels. While the correct, comfortable, and beautiful performance of a piece is supported by technical skills, the aesthetically pleasing and enjoyable performance of the piece is achieved through interpretive studies. On the other hand, the main goal in vocal training is the correct, beautiful, and effective interpretation of works with an aesthetic understanding.

Aesthetics and technique are two inseparable and essential elements of art. Aesthetics is associated with beauty, without which art cannot truly exist or be fully appreciated. Technique, on the other hand, refers to the skillful execution of an activity in the most effective manner and represents the realization and application of theoretical knowledge (Kaygısız, 2017). Like all art forms, the art of singing possesses both technical and aesthetic dimensions. In order to achieve aesthetic excellence, a singer must overcome technical challenges (Marchesi, 1986). However, this should not mean neglecting the aesthetic aspect of the art of singing by giving weight to its scientific and technical aspects. Approaching singing not solely from a scientific or technical perspective, but also as an aesthetic art from a philosophical standpoint, allows for a deeper understanding of the depth and sublimity of the art of singing. Such an approach ensures that performance skills become more accurate, expressive, and effective, as well as more balanced, stable, and consistent. In this context, vocal training should address both the technical and aesthetic aspects of the art of singing, offering an experience in which the three fundamental dimensions of human existence—mind, body, and soul (emotion)—function in complete harmony. Ultimately, efforts should be made to interpret musical works through an experiential process that fully reflects the emotions of the soul.

Webb (1946) states: “The truth is, the better the performance, the more relaxed, captivated and distanced from criticism the listener becomes. Good interpretation is certainly not something very conspicuous… I think most people would agree that the essence of interpretation is the ability to quickly grasp the essence and content of a song and to convey its message clearly and precisely… Interpretation begins with capturing the spirit of the song; whether this spirit is expressed in the music or the words, once this fundamental importance is grasped, there is no escaping it. This approach to interpretation identifies the singer with the song in such a way that it confines him to his own little world.” According to Webb (1946), the singer sometimes becomes the narrator of an exciting

or interesting story, and sometimes sings about himself, expressing his own thoughts and feelings. Of course, it is a fact that, with regard to various song genres, no artist will use the same colors or apply them in the same way.

According to Hemsley (1998), atmosphere in a song can be considered objectively in terms of the mood created in the singer by a scene from a musical, opera, or musical film, or by the situation that inspired the music and text. Accordingly, for example, a crimson sunset does not have an atmosphere/meaning in itself. This comes from the subjective reaction of the poet and composer. In this context, singers, as a requirement for a good interpretation of a song, must recreate the situation that gave rise to the poem and music in their imaginations, react to this situation and thus convey the atmosphere by singing.

The sixth sub-problem of the research, "What is the importance and role of choosing the right repertoire in developing the singing skills of prospective music teachers?", indicates that choosing the right repertoire plays a significant role in improving singing skills. Choosing the right repertoire is primarily important for protecting vocal health and using the voice correctly, comfortably, and effectively. Furthermore, it is crucial for interpreting a piece correctly, beautifully, and effectively from a technical and musical perspective, and for doing justice to the composer and the work. A good repertoire will enable students to recognize their own vocal characteristics and come to class more willingly. Therefore, choosing the wrong piece can both discourage a student from the class and seriously damage their voice, resulting in a below-average performance. In this context, choosing a piece that suits one's taste and talent will allow them to interpret the piece comfortably and confidently, and with enjoyment and willingness. Because pieces that allow a person to showcase their best talent are important for both the singer and the listener.

Repertoire is a critical component of the teaching process at all levels of music education, and there is a need for further examination of repertoire selection practices for vocal training, particularly by music educators at the university level. The repertoire that music educators choose for their students is essentially the curriculum. A musical work rich in concepts encompassing musicality, performance, history, cultural awareness, and aesthetics is a tool that enables students to acquire these skills and meets the instructors' goals (Forbes, 2001). According to Luckstone (1948), instructors should be "competent and experienced" in this regard (cited in Stephenson, 2013).

Young students in the initial stages of vocal training are at a critical point in terms of musical and physiological development, and repertoire plays a crucial role in this development. If the repertoire is chosen correctly at this stage to capture the student's musical interest, the student can safely develop their voice, musicality, and technical skills. If the repertoire selection is incorrect, vocal health will be jeopardized, and their love of music and musical enthusiasm may be irreversibly damaged. In this context, the repertoire should deeply engage the student's interest in music, earning their appreciation and admiration, for further practice and study. Songs, because they are meaningful and connected to emotions and intellect, allow the student to focus more on the fundamental processes of the voice (posture, mental-spiritual and physical comfort, breathing, producing the voice in the correct position, etc.) (Jung, 2010).

According to Rock (2005), each student is different in terms of personal and musical characteristics, and therefore repertoire selection is largely subjective. Thus, the goal of a vocal coach should be to enable the student to understand and learn how to use their own voice to develop their natural vocal ability. In this context, vocal coaches must know each student's vocal characteristics, age, vocal maturity level, needs, developmental level, talent level, musical capacity, and language skills in order to select literature suitable for their voice. It is also important to know the student's personality and personal preferences regarding the emotional expression required by the music. Finally, the selected literature should not hinder the technical developments gained through vocal training.

The seventh sub-problem of the research, "What is the importance and role of technological advancements in improving the singing skills of prospective music teachers?", indicates that technological advancements play a significant and effective role in improving singing skills. With technological advancements, access to many works, study techniques, and theoretical information has become easier and faster. Technology allows access to the sheet music of musical works, and not only that, but also to see performances in different styles, listen to examples of works from many master artists and develop critical skills. Thanks to online lessons, students can improve their voice and singing skills with an instructor. Sound recording devices, computer programs, and various applications offer the opportunity to record and listen to the sound, analyze it, and identify errors, as well as providing opportunities for both analysis and evaluation by improving the voice. Digital applications allow for regular practice, working on rhythm and intonation by singing with digital platforms, and singing along to songs, as many songs also have karaoke versions. Thus, it's possible to practice with accompaniment at any time and in any key, without needing to find an accompanist, thanks to applications that allow for key changes. Furthermore,

students can immediately access translations of works in foreign languages, easily learn the correct pronunciation and interpretation of the piece through video recordings, and improve their own interpretation. Today, technological applications also offer easy access to events such as operas, concerts, and recitals, providing access to events not staged in our country, as well as access to a wealth of useful information about opera and vocal training, allowing students to follow and learn from those involved in this field.

21st-century educational trends are replacing traditional models that prioritize rote learning and memorization with models that cultivate students who can think critically and solve problems. Today, students have high expectations regarding their ability to access, evaluate, analyze, and synthesize information. Technology, which supports the application of active learning models, is adaptable to different learning styles and provides motivation for students with specific learning needs. Technology, now integrated into curricula at all levels of education, offers new opportunities in many areas for both teachers and students. Individuals are increasingly using technology to acquire knowledge through their own experiences and discoveries. They can then apply these newly developed thought systems to their personal lives and subsequently to their professional lives.

Traditional music education has become highly effective and multifaceted in many countries around the world with the support of technology. Technologies included in the curricula of higher education institutions providing professional music education are transforming the music learning environment into a technological learning center. Technologies such as MIDI, computers, software, the internet, television, DVDs, CDs, CD-ROMs and electronic instruments used in these schools improve the knowledge, skills and performance of future musicians and music educators and increase their creativity and motivation (Tecimer Kasap, 2007).

According to Arapgirlioğlu (2003), technology, which brings a new perspective to the understanding of education, also provides students with a comprehensive educational opportunity in music education. Technological tools guide students in actively making music, enabling them to strengthen their creativity by creating their own original compositions with pleasure through active participation. As a result of many studies, it has been observed that thanks to the application of technology in music education, music lessons become more interesting for students, help them gain self-confidence, provide more efficient and effective learning, strengthen group work, positively affect critical thinking and problem-solving skills, enable the understanding of music in its scientific and artistic dimensions, and increase enjoyment of music lessons through active participation. This situation is extremely important in the development of students' ability to interpret musical works. Thus, while technology adds an active participation and intellectual dimension to music, it also significantly contributes to the development of students' interpretation skills.

McCoy (2014) notes that technology affects singing in many ways beyond voice analysis. According to the author, aspiring singers heavily utilize internet resources to promote themselves and advance their careers, including social networks and personal websites. Establishing a group on Facebook is often the most effective and efficient way for vocal coaches and choir/orchestra conductors to connect with their studios or communities. Furthermore, an increasing number of singing teachers are offering lessons via video conferencing over the internet.

Recommendations

The following recommendations are presented based on the results of the research:

1. For prospective music teachers, using their voice naturally and healthily, as well as correctly, beautifully and effectively, is crucial for becoming more knowledgeable, confident and successful, both during their undergraduate education and in their future professional lives. In this context, vocal training courses should be conducted with this understanding and emphasis should be placed on developing students' singing skills.

2. In developing the singing skills of prospective music teachers, theoretical knowledge related to vocal training (human voice, its characteristics, voice types, voice production processes “breathing, phonation, resonance, articulation”, etc.) should be emphasized and included in the lessons.

3. In vocal training classes, sufficient emphasis should be placed on the development of technical knowledge and practical skills at the application level. These include correct posture; mental, emotional and physical relaxation; proper breathing techniques and breath control; establishing an effective voice–breath connection; and producing a healthy and accurate vocal tone. In addition, instruction should address resonance, register transitions (head voice, middle range, and chest voice) and clear articulation. Students should also be made fully aware of the fundamental importance of these components in achieving sustainable and effective vocal performance.

4. Providing information on the theoretical aspects of musical works—such as the composer, historical period, genre, performance style, the meaning of the lyrics in the case of foreign-language works, as well as the subject

matter and narrative—lays a more solid foundation for engaging with the pieces and, in turn, enhances students' singing skills.

5. Emphasis should be placed on speaking exercises, especially for songs written in foreign languages, with diction, pronunciation, and articulation exercises. This work will show respect and care for the language in which the work was written, and the performance will be more accurate and aesthetic.

6. Emphasis should be placed on the accurate, beautiful, and effective interpretation of works, as well as on theoretical and technical studies, and due diligence should be given to efforts aimed at this goal. In addition, it would be beneficial to encourage students to participate in scientific activities, read publications on the human voice and its training, listen to different musical and singing styles, attend concerts, and engage in similar activities.

7. In individual voice training lessons, careful consideration should be given to selecting repertoire that is appropriate for each student and their specific needs. This is essential not only for the student's vocal health and technical and musical development but also for fostering their interest, curiosity, motivation and engagement in the lesson, ultimately ensuring more effective and high-quality instruction.

8. The importance of technology in education today is undeniable. In this context, students should be guided on the significance and proper use of technology in developing their singing skills, and technological applications should be appropriately integrated into lessons.

REFERENCES

Appleman, D. R. (1986). The Science of Vocal Pedagogy. Bloomington: Indiana University Press.
Arapgirlioğlu, H. (2003). Müzik teknolojisi ve yeni yüzyılda müzik eğitimi [Bildiri sunumu]. Cumhuriyetimizin 80. Yılında Müzik Sempozyumu, 30-31 Ekim, İnönü Üniversitesi Eğitim Fakültesi, Malatya.
Arlı, M. ve Nazik, M. H. (2014). Bilimsel Araştırmaya Giriş. Ankara: Gazi Kitabevi.
Balcı, A. (2004). Sosyal Bilimlerde Araştırma-Yöntem, Teknik ve İlkeler. Ankara: Pegem A Yayıncılık.
Bicknell, J. (2015). Philosophy of Song and Singing. London and Newyork: Taylor & Francis.
Buchanan, H. J. (2014). Body Mapping: Enhancing Voice Performance Through Somatic Pedagogy. Teaching Singing in the 21st Century. (Editors: Scott D. Harrison and Jessica O'Bryan), Springer.
Davran, Y. (1997). Şarkı Söyleme Sanatının Öyküsü. Ankara: Önder Matbaacılık.
Dayme, M. B. (2009). The Role and Function of The Voice" in Dynamics of the Singing Voice. 5th ed. 43-55. Springer.
Demirel, Ö. (1999). Kuramdan Uygulamaya Eğitimde Program Geliştirme. Ankara: Pegem A Yayıncılık.
Ekici, T. (2020). Ses Eğitiminde Repertuvar. Eğitim Bilimleri-Teori, Güncel Araştırmalar ve Yeni Eğilimler (Edt. Harun Şahin). IVPE.
Ekiz, D. (2003). Eğitimde Araştırma Yöntem ve Metodlarına Giriş. Ankara: Anı Yayıncılık.
Harrison, S. D. ve O'Bryan, J. (2014). The Future of Singing Pedagogy. Teaching Singing in the 21st Century (Editors: Scott D. Harrison and Jessica O'Bryan), Springer.
Hemsley, T. (1998). Singing and Imagination: A Human Approach to a Great Musical Tradition. Oxford: Oxford University Press.
Howard, V. A. (1982). Artistry: The Work of Artists. Indianapolis: Hackett Publishing Co.
Jung, J. Y. (2010). Young singer and the repertoire. (Doctoral dissertation). University of Washington, USA.
Kagen, S. (1960). On Studying Singing. Newyork: Dover Publications, Inc.
Kaygısız, M. (2017). Müzik Tarihi- Başlangıcından Günümüze Müziğin Evrimi. İstanbul: Kategori Yayıncılık.
Khan, S. I. (1994). İnsan ve Evren Arasındaki Köprü, Müzik. (çev. Kaan Ökten, Tuğrul Ökten). İstanbul: Arıtan Yayınevi.
Marchesi, M. (1986). Marchesi's Vocal Method. Edwin Ashdown Limited, London, NW2 6QR.
McCoy, S. (2014). Singing Pedagogy in the Twenty-First Century: A Look Toward the Future. Teaching Singing in the 21st Century (Editors: Scott D. Harrison and Jessica O'Bryan), Springer.
McKinney, J.C. (1994). The Diagnosis and Correction of Vocal Faults. Genevox Music Group, U.S.A.
Mingle, A. J. (2018). Applications of somatic education principles to voice pedagogy. (Doctoral dissertation). State University of New Jersey.
Morrison, A. (2006). Singing Lesson. Academic Search Complete, 14 (5), 11-14.
O'Bryan, J. ve Harrison, S. D. (2014). Positioning Singing Pedagogy in the Twenty-First Century. Teaching Singing in the 21st Century (Editors: Scott D. Harrison and Jessica O'Bryan), Springer.
Potter, J. ve Sorrell, N. (2014). A History of Singing. Cambridge University Press. UK
Rao, D. (1988). Craft, singing craft and musical experience: A philosophical study with implications for vocal

music education as aesthetic education. (Doctoral dissertation). Northwestern University.
Rock, C. (2005). The Application of Vocal Literature in the Correction of Vocal Faults. University of Connecticut.
Rogers, C. K. (2013). The Philosophy of Singing. London: Whitaker Press.
Sabar, G. (2008). Sesimiz "Eğitimi ve Korunması" İstanbul: Pan Yayıncılık.
Santelli, C. E. (2023). Reasoning the voice: Toward a diagnostic and prescriptive technique in the teaching of singing. (Doctoral dissertation). Louisiana State University.
Serra-Dawa, S. (2014). The Teacher-Student Relationship in One-to-One Singing Lessons: An Investigation of Personality and Adult Attachment. Teaching Singing in the 21st Century (Editors: Scott D. Harrison ve Jessica O'Bryan), Springer.
Sparshott, F. E. (1982). The Theory of the Arts. Princeton: Princeton University Press,
Stark, J. (2003). Bel Canto: A History of Vocal Pedagogy. United Kingdom: University of Toronto Press.
Stephenson, D. G. (2013). An investigation of selected collegiate voice teachers' descriptions of repertoire selection practices. (Doctoral dissertation). University of South Carolina.
Tecimer Kasap, B. (2007). Müzik eğitiminde teknolojik yaklaşımlar. (Tam Metin Bildiri). Türkiye Cumhuriyeti Başbakanlık Atatürk Kültür, Dil ve Tarih Yüksek Kurumu Başkanlığı, 38. Uluslararası Asya ve Kuzey Afrika Çalışmaları Kongresi (ICANAS), 10-15 Eylül, Ankara, Türkiye.
Vennard, W. (1967). Singing: The Mechanism and The Technic. Newyork: Carl Fischer, Inc.
Webb, D. (1946). Some Aspects of Song Interpretation. 72nd Sess./ pp. 123-139. Taylor & Francis, Ltd. on behalf of the Royal Musical Association.
Yıldırım, A. ve Şimşek, H. (2004). Sosyal Bilimlerde Nitel Araştırma Yöntemleri. Ankara: Seçkin Yayıncılık.
Yiğit, F. (2000). Müzik Eğitiminde Kodaly Metodunun Rolü. (Yüksek Lisans Tezi). Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü, İzmir.
Yin, R. K. (1984). Case Study Research: Design and Methods. Beverly Hills, CA: Sage.

AI Prompt Writing Rubric: A Validity and Reliability Study

Nurcan İnan
Marmara University, Educational Sciences ,Department Of Curriculum And Instruction, İstanbul, Türkiye
nrcinan@gmail.com
ORCİD: 0000-0003-3612-6011

Sibel Cengizhan
Marmara University, Educational Sciences ,Department Of Curriculum And Instruction, İstanbul, Türkiye
sibel@marmara.edu.tr
ORCİD:0000-0001-5862-2927

Seyfi Kenan
Marmara University, Educational Sciences ,Department Of Curriculum And Instruction, İstanbul, Türkiye
seyfi.kenan@marmara.edu.tr
ORCİD:0000-0002-3773-7693

ABSTRACT
This study introduces the development and validation of an analytical rubric designed to teach sixth-grade students how to write effective prompts. An initial draft rubric was developed based on a literature review on ChatGPT and prompt engineering, as well as the opinions of seven experts. The rubric was piloted with 32 sixth-grade students. We re-evaluated content validity, assessed construct validity through factor analysis, and measured internal consistency through Cronbach's alpha. During validation, four items were removed due to low common variance, and item 10 was excluded for redundancy. The final version demonstrated robust construct validity and internal consistency. Moreover, the Fleiss' kappa value of 0.29 showed fair to moderate interrater agreement. Implications for Practice or Policy: This section presents implications for educators, policymakers, students, and researchers: (1) Policymakers can create assessment tools aligned with AI-integrated curricula, using the developed rubric as a guide. (2) Educators can use the rubric for lesson planning, assessing prior knowledge, and measuring skill development. (3) Researchers may build foundational K-12 assessment studies based on this work. (4) Students can enhance their AI communication by writing clearer, more polite, and purposeful prompts, thereby improving their written expression and self-assessment skills.
Keywords: Artificial intelligence prompt, ChatGBT 3.5, K-12, Validity and reliability, Writing rubric

INTRODUCTION

The concept of intelligence

Intelligence has long been regarded as one of the core concepts that has captivated philosophers, psychologists, and scientists from Ancient Greece to the present day (Sternberg, 2005). Despite its enduring significance, there is still no consensus among researchers regarding the nature and scope of intelligence (Woodcock, R. W.,1990; Solso, 1995; Halonen & Santrock, 1996). While intelligence is often described as one's ability to adapt to their environment, solve problems, and learn from experience, how these processes function across different contexts remains a subject of ongoing debate (Sternberg, 2005). In this debate, various theorists proposed differing classifications of the core abilities that constitute intelligence. Spearman's Theory of General Intelligence, for example, claims that intelligence is a unified construct, with mental energy serving as the driving force behind all cognitive actions (Köksal, 2007). In contrast, Thorndike's (1920) Multifactor Theory conceptualizes intelligence as the capacity to effectively navigate novel situations and respond with appropriate solutions. Similarly, Sternberg's (1985) Triarchic Theory of Intelligence divides intelligence into three interrelated dimensions: analytical, creative, and practical. These frameworks offer more comprehensive perspectives on human intelligence and contribute significantly to our understanding of cognitive processes.

Research on intelligence has consistently highlighted problem-solving, decision-making, and environmental adaptability as core components of the construct (Sutarso, 1998; Budak, 2000; Rau, 2001). These components suggest that intelligence is not solely rooted in cognitive processes but is also closely linked to physical and emotional domains (Damasio, 1999). In this context, Piaget defined intelligence as one's ability to adapt to their environment and organize their thoughts and behaviors accordingly (Clark, 2019). By the 1980s, intelligence began to be understood as a measurable construct, often expressed through intelligence quotient (IQ) scores (Hoerr, 2000). Alfred Binet laid the foundation for the IQ concept by developing the first systematic intelligence test aimed at determining children's mental age (Myers, 1998). However, subsequent research demonstrated that IQ tests fall short of capturing the full scope of an individual's cognitive potential, emphasizing the need to

conceptualize intelligence as a multidimensional construct (Riggio et al., 2002). Similarly, Wechsler (1943) argued that intelligence assessments should encompass not only cognitive but also emotional and social components. Supporting this perspective, Gardner (1983; 1999) proposed the theory of multiple intelligences, contending that intelligence extends beyond mathematical and linguistic abilities and that individuals may exhibit exceptional strengths across various domains. These diverse efforts to understand human intelligence have laid the groundwork for developing models that mimic cognitive processes, ultimately enhancing machines' capacity to learn, solve problems, and adapt to their environments (McCarthy, 2007). Furthermore, these theoretical advancements have played a foundational role in shaping the field of artificial intelligence (AI). Notably, McCarthy (2007) redefined intelligence not as an exclusively human trait but as a phenomenon observable in some animals and even certain machines, thereby broadening the definition to include artificial entities.

Artificial intelligence

The concept of machine intelligence first entered academic discourse in 1950 through Alan Turing's seminal question, "Can machines think?", which laid the foundation for what became known as the Turing Test. This test was designed to evaluate whether a machine could exhibit human-like cognitive abilities. In the test, a human evaluator engaged in written communication with two participants– one human and one machine. If the evaluator could not reliably distinguish the responses from the machine from those from the human, the machine was considered to have the ability to think (Turing, 1950). Turing's work provided a theoretical framework that not only established the foundation of AI but also guided the development of modern AI systems. Over time, this framework has spurred practical applications of AI across a wide range of fields, including education, healthcare, finance, agriculture, industry, retail, security, transportation, logistics, law, and the creative industries. For example, an AI-based portable electronic assistive device was developed to support visually impaired individuals in navigating their environments independently (Shariff, 2020). Similarly, AI has been employed to detect safety and quality issues, while it has been used to automate personalized advertisement targeting (Davenport & Ronanki, 2021). In the field of education, particularly in language learning, AI applications have been developed for speech and pronunciation recognition, as well as for answering student queries, thus demonstrating the potential to support personalized learning processes (Hill et al., 2015). These diverse implementations have led to varying definitions of AI, tailored to the specific needs and perspectives of each discipline. Computer engineers, for instance, describe AI as a form of machine learning that utilizes artificial neural networks to apply algorithms to data for pattern recognition, decision-making, and predictive analysis (Say, 2018; Marr, 2020; Chivers, 2020; Yılmaz, 2022). In the field of medicine, scholars conceptualize AI within the framework of the "artificial human" (Aydın & Değirmenci, 2018), while philosophers explore its implications for consciousness, mind, and reasoning (Köse, 2022). Education researchers, on the other hand, describe AI in terms of its parallels with cognitive processes, focusing on a computer's ability to reason, solve problems, generalize, adapt, comprehend language, and make decisions in ways that resemble human cognition (Shidiq, 2023).

The combination of high computational power, vast data volumes, and advanced machine learning algorithms has recently driven significant progress in AI-based technologies (Russell, 2021). Thus, the capabilities of artificial intelligence now extend far beyond language processing and decision-making; they encompass a broader spectrum that includes visual perception, learning, autonomous action, and even creative thinking (Altıntop, 2023). In parallel with this expansion, numerous models have emerged, many of which are trained using complex architectures such as artificial neural networks developed through deep learning techniques (Goodfellow et al., 2016). Among the most notable are large language models (LLMs) built on transformer-based architectures (Sutton& Baro, 2018), which have become closely associated with AI in the public imagination, evident in the widespread use of voice and text assistants (e.g., Siri, Alexa, and ChatGPT). Particularly, generative AI tools, including language models and visual content generators, have enhanced users' productivity and enabled them to perform various tasks more efficiently (Kutlucan & Seferoğlu, 2024). One such tool is Chatbot Generative Pre-Trained Transformer (ChatGPT), released by OpenAI at the end of 2022 (ExcelinEd, 2023). Free access and its ability to produce highly relevant responses to user prompts (Günbatar & Ağgün, 2024) have underscored the growing necessity for clear and effective communication in written interaction with AI. This form of communication is defined by the concept of a "prompt," which refers to the initial input text designed to elicit a specific response from a language model (Bea et al., 2024). In other words, a prompt functions as a guiding input that steers the output toward a particular task (Brown et al., 2020). In this regard, the technique known as "prompt engineering" assumes a key role in optimizing the performance of AI systems.

Prompt engineering refers to the strategic formulation of natural language inputs and the refinement of interactions with LLMs. Prompts—consisting of task instructions, input data, and the expected output format

serve as a critical interface between human intent and machine response during the inference phase. Among the overall prompt strategies outlined by OpenAI, several stand out: articulating tasks and expected outputs with precision, supplying reference texts to minimize hallucinated content, decomposing complex assignments into manageable components, and allowing the model time to engage in reasoning processes (OpenAI, n.d.). In this light, the crafting of effective prompts is increasingly recognized as a pivotal factor influencing both the relevance and accuracy of LLM-generated outputs. Systematic prompt design plays a crucial role in optimizing model performance and tailoring outputs to specific needs (Mesko, 2023; Chakraborty et al., 2024; Schulhoff et al., 2024). Moreover, prompt strategies may be categorized based on their underlying objectives—ranging from guiding model behavior without the need for retraining to fostering reasoning capabilities and mitigating the risk of misinformation (Chakraborty et al., 2024). Anchored in these strategic insights, we developed a series of prompt steps and identified related behavioral subcomponents. Accordingly, the prompt dimensions and their corresponding behavioral indicators were structured as follows:

1.Clarity and Precision in Task Definition (a) Clearly articulates their requests when defining the task to a language-based AI. (b)Breaks down the task into small, manageable components as concise sentences.
2. Profile Creation (a) Creates a level-appropriate profile to guide the AI's responses.
3. Grammar Usage and Expression (a) Adheres to grammatical rules. (b) Provides accurate commands when defining the task for the language-based AI. (c) Enhances the effectiveness of the prompt by using adjectives, conjunctions, and adverbs correctly.
4. Politeness and Professional Tone (a) Employs polite expressions when interacting with the language-based AI. (b) Adopts a tone consistent with the defined task and user profile.
5. Output Refinement and Attention to Detail (a) Evaluates the generated output. (b) Writes follow-up prompts with refined details to complete or improve nuanced aspects of the output.
6. Critical Evaluation and Verification of Information (a) Critically evaluates the accuracy and reliability of the information provided by the language-based AI.

Measuring these dimensions and behavioral indicators, alongside evaluating students' academic performance, plays a critical role in analyzing the effectiveness of instructional processes (Ministry of National Education [MoNE], 2024). In this regard, the development of a valid and reliable tool for assessing prompt writing skills is considered essential. Guided by this need, the present study aims to develop a rubric capable of validly and reliably assessing the prompt writing competencies of sixth-grade students.

Rubrics, recognized as a specialized form of checklist, serve not only to articulate the ideal qualities of a completed task to students but also to provide teachers with a structured mechanism for assessing and scoring student performance in detail. Rubrics outline a set of specific criteria, each accompanied by clearly defined performance levels, which makes them effective tools not only for evaluating performance but also for tracking student progress and guiding learning (Brookhart, 2013). With access to these scoring criteria, students are better equipped to understand the expectations for improving their future work (Moskal & Leydens, 2000). In addition, rubrics enable individuals to assess their own performance (Mertler, 2001; Oakleaf, 2009), while offering both teachers and students constructive feedback on their strengths and weaknesses (Hall & Salmon, 2003). Teachers can also systematically monitor learners' progress toward achieving educational goals through rubric-based assessment (Arter, 2002).

Rubrics are often classified into two main categories: holistic and analytic. While holistic rubrics provide an overall assessment of a student's work (Taggart et al., 1998), analytic rubrics examine student performance in detail, aligned with specific learning objectives, and are regarded as one of the most effective and widely used assessment tools (Reeves, 2011). Due to their structured and targeted nature, analytic rubrics are often seen as more practical and comprehensive than holistic ones (Amanvermez İncirkuş & Beyreli, 2019). In this sense, we adopted an analytic rubric in this study to assess sixth-grade students' written communication skills using language-based AI tools, specifically ChatGPT. In developing the rubric, we followed the steps outlined by Goodrich Andrade (2001), including identifying the assessment criteria, defining the levels of performance, and consulting expert opinions to ensure the validity and clarity of the rubric. Then, we performed a comprehensive analysis to assess the validity and reliability of this rubric.

Literature review

In the relevant literature, a plethora of studies focus on forming a solid theoretical foundation for analysis by scrutinizing the role of ChatGPT and similar AI tools in education, the significance of prompt instruction, the function of rubrics in educational settings, and the necessity of evaluating prompts. In the realm of education, a growing body of research has emerged regarding the integration of language-based AI tools, with increasing attention being paid to the pedagogical applications of ChatGPT. Recent studies have explored the multifaceted role of ChatGPT in instructional contexts. For example, Qureshi (2023) noted that universities attempt to

leverage technology to enhance student learning and that instructors serve as facilitators in this process. Similarly, Mhlanga (2023) reported that ChatGPT is currently under examination for a variety of academic purposes, including instruction delivery, language learning, educational feedback, and assessment. Advances in AI models enable the development of innovative educational applications capable of fundamentally transforming learning experiences. Chiu et al. (2024) examined the educational potential of ChatGPT and other AI technologies, highlighting their capacity to personalize student learning and enhance engagement. Moreover, AI tools support teachers in tailoring lessons to individual needs and hold promises for fostering personalized learning experiences in university settings (Vatansever, 2024). Xu (2022) further emphasized that AI can serve as a valuable aid in problem-solving and rapid-response scenarios, particularly within STEM (Science, Technology, Engineering, Mathematics) education. In their 2024 bibliometric analysis on ChatGPT-based learning, Ching-Yi Chang and colleagues found that its use was most concentrated in educational technology, English language learning, and STEAM (Science, Technology, Engineering, Art & Mathematics)-related studies.

Previous research on prompt engineering indicated that the quality of prompts significantly influences the quality of output generated by LLMs for relevant tasks (Knoth et al., 2024). In their study exploring the relation between AI literacy and prompt engineering skills among university students, Knoth et al. (2024) presented experimental evidence showing that advanced prompt engineering fosters the production of higher-quality LLM outputs and enhances users' ability to harness the potential of such technologies more effectively. Prompt writing is positioned as a measurable skill that distinguishes individuals who can productively utilize LLMs from those who struggle to generate desired outcomes. Another study highlighted that while individuals without expertise in AI can engage in prompt engineering, their limited understanding of LLM capabilities and the tendency to mimic human-human instructions hinder systematic progress (Zamfirescu-Pereira et al., 2023). Similarly, relevant research showed that students often approach LLM-based AI systems as though interacting with a human, using socially desirable phrases such as 'Hello!' and 'Thank you.' This behavior is attributed to the human-like interfaces and conversational abilities of LLMs, which lead users to anthropomorphize these systems (Bewersdorff et al., 2025). Collectively, the previous research underscores the significance of equipping individuals, particularly K-12 students, with prompt engineering skills.

Although several studies examined the educational applications of ChatGPT at the K-12 level, research specifically focused on assessing students' prompt writing skills remains notably scarce. Among the limited body of work, Kobara et al. (2024) developed analytic rubrics to assess learning outcomes within AI education programs for K-12 students and analyzed the reliability of these rubrics through an AI-focused educational workshop. The findings revealed that, while the rubric demonstrated internal consistency, its inter-rater reliability remained poor. The study highlights the need for more tailored rubrics and the identification of appropriate methods for assessing learning outcomes.

Ultimately, the analytic rubric developed in the present study (Appendix A) is expected to contribute to the evaluation of prompt writing skills at the K-12 level. Moreover, we anticipate that while both researchers and practitioners can reap such a context-specific analytic assessment tool, our findings will guide future rubric development efforts in this emerging area.

THE STUDY

In this study, we employed a methodological design to ensure the validity and reliability of the rubric developed. During the 2023–2024 academic year, we administered the draft version of the rubric to 32 sixth-grade students attending a private school located on the European side of Istanbul. Two teachers and one researcher then evaluated the rubric. In the development and validation phases of the rubric that assesses the commands from students during their interactions with ChatGPT 3.5 and their prompt writing steps, we adhered to the following steps (Taggart et al., 1998; Brookhart, 2013; Schoepp et al., 2018; Amanvermez İncirkuş & Beyreli, 2019; Bhatnagar et al., 2021):

Reviewing the literature on prompt engineering and ChatGPT

Initially, we identified the key components of prompt writing based on the existing literature on prompt engineering and ChatGPT (Spasić & Janković, 2023; Vairamani & Nayyar, 2024; Öz, 2024; Lindley & Whitham, 2024; Li & Klabjan, 2024; Ein-Dor et al., 2024; OpenAI, n.d.; Prompting Guide, 2025). In AI-mediated writing tasks, the process of generating commands is grounded in a conceptual formula that consists of three main elements: the task, the instructions, and the role (John, 2023). The 'task' refers to clearly and explicitly stating what the prompt aims to achieve. In the pilot study, an example of such a task was having students instruct a language-based AI model to compose a poem. 'Instructions' denote the steps the AI model is expected to follow in executing the task, while the 'role' defines the persona/profile the model should adopt

while generating the text (Melanson & Maman, 2024). In this study, we noted that students asked ChatGPT to adopt the perspective of a sixth-grade student and to produce poems suited to that level. Accordingly, as prompt writing skills need to be designed in alignment with students' language and written communication abilities, we identified appropriate criteria and tasks for the assessment process.

Creating a draft rubric

In constructing the dimensions of the prompt rubric, we utilized both native language learning outcomes and Information and Communication Technologies (ICT) curriculum objectives as foundational criteria, as these are considered prerequisites for the language and written communication skills underpinning prompt writing steps. We tried to ensure that the tasks expected to be carried out using AI tools for text generation were aligned with the topics covered in class. Accordingly, key aspects of native language competencies (e.g., grammar, spelling conventions, and expression) were incorporated into the rubric as sub-dimensions. In addition, we included complete and error-free writing, as well as the use of polite language, in the draft rubric as sub-dimensions and observable behaviors to be assessed (Oz, 2023). The draft rubric consists of 7 dimensions and includes 11 behavioral indicators evaluated across four performance levels: Excellent, Proficient ,Basic, and Needs Improvement. We structured the drafting process of the rubric around the following steps to ensure its validity and reliability.

FINDINGS

Calculating content validity rates and content validity indices through expert opinions

To ensure the content validity of the draft rubric, we refined the items based on feedback from seven experts employed in various schools and universities. These experts included a computer science teacher, a native language teacher, an academic specializing in computer science, two curriculum and instruction specialists, and two experts in educational measurement and evaluation. In this process, we employed the Lawshe technique. According to Lawshe (1975), the minimum content validity ratio (CVR) value for an instrument evaluated by seven experts should be 0.99. Accordingly, we calculated the CVR value to be 1.00 for our draft rubric and made several revisions to some rubric items in line with expert opinions. For example, the indicator initially worded as "Defines the task to ChatGPT clearly and explicitly" was revised by one expert as "Clearly articulates their requests while defining the task to a language-based AI tool." Another expert recommended including adverbs of frequency in the indicator (e.g., always, usually). Moreover, we separated items involving multiple actions (e.g., "Always breaks down the task into small, manageable components in concise sentences and creates an appropriate profile to complete the task") into distinct behaviors based on expert opinions, resulting in an increased number of indicators. In addition, descriptions under the "Grammar Usage" and "Output Refinement and Attention to Detail" dimensions were rewritten to ensure clarity and precision.

Following these revisions, the total number of behavioral indicators increased to 14. We also restructured the performance levels based on expert recommendations and finalized them as "Excellent," "Proficient," "Partially Proficient," and "Not Proficient." The findings related to Lawshe's analysis are presented in Table 1.

Table 1. Findings of the CVR values

	Relevant	Useful/Need to be revised	Irrelevant	Total number of experts (N)	Responses to "Relevant" Items (NG)	CVR	Decision
Item 1	4	3	0	7	7	+1.00	Accept
Item 2	5	2	0	7	7	+1.00	Accept
Item 3	6	1	0	7	7	+1.00	Accept
Item 4	5	2	0	7	7	+1.00	Accept
Item 5	6	1	0	7	7	+1.00	Accept
Item 6	7	0	0	7	7	+1.00	Accept
Item 7	5	2	0	7	7	+1.00	Accept
Item 8	5	2	0	7	7	+1.00	Accept
Item 9	6	1	0	7	7	+1.00	Accept
Item 10	6	1	0	7	7	+1.00	Accept
Item 11	7	0	0	7	7	+1.00	Accept

Pilot study and administering the rubric

We carried out the pilot study for the draft rubric with the designated sample. A total of 32 sixth-grade students received a brief introduction to ChatGPT in their ICT class and were subsequently asked to complete a poetry-writing task using the platform. To ensure that participants were adequately prepared to complete the task, we

administered a Turkish language test to assess their basic language skills before the pilot study. This test helped identify deficiencies in students' language abilities. After providing the necessary support to address these gaps, we conducted the pilot study to further test the validity and reliability of the rubric. Prior to implementation, ethical approval forms were obtained from the Ministry of National Education, the participating school, the affiliated university, and the students' parents.

In the pilot study, students were allowed to complete the task based on their current written communication skills, without prior instruction on how to interact effectively with a language-based AI tool. The poetry-writing task allowed for natural language use; therefore, students could engage authentically with the AI. The evaluation involved three raters: the researcher, a Turkish language teacher, and an ICT teacher. Each student was assigned a single task, resulting in a total of 32 rubric-based assessments. Following the evaluation, we received no revision suggestions from the raters.

Establishing the construct validity of the rubric

We examined the clustering of behavioral indicators in the rubric using factor analysis. Specifically, we employed an exploratory factor analysis (EFA) with principal component analysis and orthogonal rotation (varimax) to determine the construct validity of this multi-dimensional scoring rubric and to uncover its factorial structure. We chose the principal component analysis as it is one of the most commonly used and practical methods in practice, while the varimax rotation was selected based on the assumption that there would be no correlation between factors.

There were no missing values in the dataset. We also calculated z-scores to detect outliers, and data within the range of $-3 < z < 3$ were retained for analysis. Among the 210 cases, we excluded nine outliers exceeding this range. In addition, the following four items with communalities of 0.5 or below were excluded from the analysis: "Providing examples in the prompt," "Specifying key terms (e.g., topic or main idea) that align with the task," "High-level alignment between the output and the prompt," and "Adhering to grammar rules." To assess the suitability of the dataset for factor analysis, we examined the Kaiser-Meyer-Olkin (KMO) value and Bartlett's test of sphericity. The KMO value was found to be .740, indicating a moderate level of sampling adequacy. Bartlett's test of sphericity produced a statistically significant result, $\chi^2(45) = 1085.443$, $p < .001$, confirming that the data exhibited multivariate normality (Büyüköztürk, 2022). We also examined the scree plot to explore the dimensionality of the items analyzed through factor analysis (Figure 1).

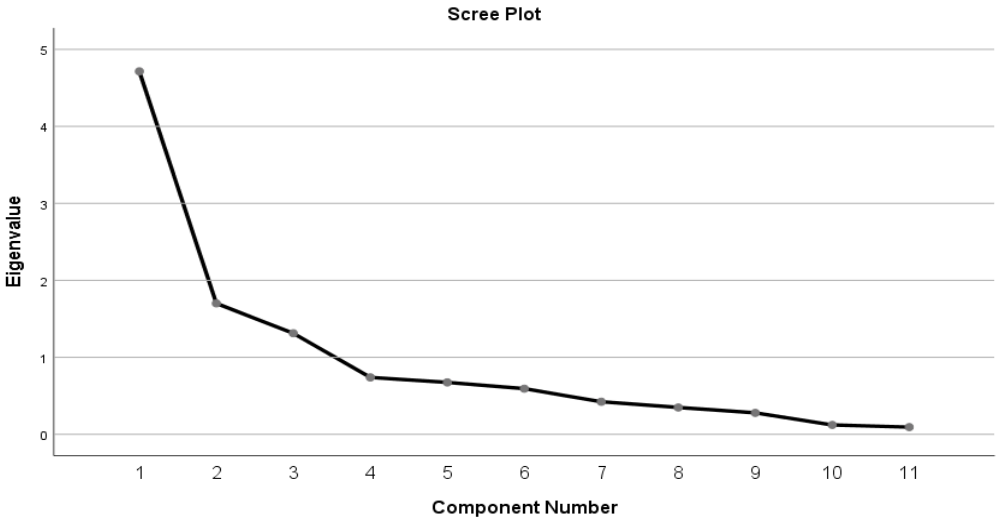


Figure 1. Scree plot

The rubric produced two factors. While the first factor accounted for 42.74% of the total variance, the second factor explained 17.08%. Table 2 presents the distribution of items across the factors along with their corresponding factor loadings.

Table 2. Factorial Structure of the Rubric, Item Factor Loadings, and Item Contributions to Common Variance

Items	Factor 1	Factor 2	Contribution to Common Variance
Item 1	0.84		0.70
Item 4	0.72		0.54
Item 9	0.88		0.78
Item 11	0.88		0.78
Item 12	0.76		0.64
Item 13	0.76		0.59
Item 6		0.89	0.79
Item 14		0.89	0.79
Item 8	0.58		0.35
Item 10		-0.98	0.84
Eigenvalue	4.26	2.54	
Variance Explained	42.74	17.08	
Total Variance Explained		59.82	

Table 2 presents the distribution of the items across two factors, along with their factor loadings, contributions to common variance, eigenvalues, and explained variance. Factor loadings indicate the strength of the relations between items and the corresponding factors. High factor loadings (.70 and above) suggest that an item is strongly represented within that factor (Deniz, 2021). Accordingly, Item 1 demonstrated a strong association with Factor 1, with a factor loading of .84, making it one of the key components of this dimension. Item 4 also loaded significantly on Factor 1 with a loading of .72. Items 9 and 11 both showed high loadings f .88 on Factor 1, further emphasizing their salience within this dimension. Items 12 and 13 lso loaded strongly on Factor 1, each with a loading of .76, reflecting their close alignment with this dimension. Items 6 and 14 both loaded on Factor 2 with a factor loading of .89, indicating their centrality in this dimension. Item 10 exhibited a negative factor loading of −.98 on Factor 2, suggesting that it represents an opposing component within this factor. Finally, Item 8 had a factor loading of .58 on Factor 1, indicating a relatively lower contribution to this factor.

In this study, only factors with eigenvalues > 1 were retained in the analysis. Accordingly, the factor analysis yielded the eigenvalues for these two factors to be 4.26 and 2.54, respectively. Factor 1 accounted for 42.74% of the total variance, and Factor 2 explained 17.08%, indicating that together, the two factors explained 59.82% of the total variance. This proportion is generally considered acceptable in factor analysis and suggests that the scale has an adequately supported construct validity (Field, 2018).

Establishing the reliability of the rubric

We calculated Cronbach's alpha coefficients to assess the internal consistency of the two dimensions of the rubric. A coefficient closer to 1.00 indicates a higher degree of internal consistency among the items (Kula & Mor, 2016). Accordingly, the coefficient was found to be .89 for the seven-item Factor 1 and .76 for the two-item Factor 2. As both values exceed the commonly accepted threshold of .70, the internal consistency of each factor can be considered adequate.

We performed an item analysis to examine the relation between Items 8 and 10. Exploring inter-item correlations basically aims to bring further evidence to the internal consistency of a measurement tool and explore the association between items (Kilmen, 2022). suggesting that participants scoring highly on one item also tended to score highly on the other. This finding also implies a genuine relation between these items rather than a random association. Given the size of the dataset (n = 201) for this analysis, this finding further contributed to the reliability of the rubric. Correlations between items are often used to assess conceptual overlap; while excessively high correlations (e.g., r > .80) may indicate redundancy, values between .40 and .70 are generally considered acceptable in scale development research (Deniz, 2021). Hence, Item 8 was retained in the rubric despite its relatively modest contribution to Factor 1. However, due to its strong inverse loading and limited theoretical alignment, Item 10—related to the use of polite language—was excluded from the final version of the rubric.

Calculating the inter-rater reliability of the rubric

Following the examination of the internal consistency of the pilot-tested draft rubric, we evaluated its inter-rater reliability using Fleiss' kappa coefficient. Fleiss' kappa is a statistical measure used to evaluate the level of agreement among multiple raters when classifying a common set of items or individuals into specific categories (Fleiss et al., 2003). However, each rating criterion inherently requires its own sub-criteria (Bıkmaz Bilgen & Doğan, 2017). Thus, the evaluation involves a progression from numerical judgments to performance-based

interpretation, ultimately leading to the measurement of consensus. In this context, assessing the kappa coefficient serves as a response to the question: "To what extent did the raters agree?" Accordingly, we calculated the inter-rater reliability coefficient using the statsmodels.stats.interrater.fleiss_kappa() function in Python. The resulting Fleiss' kappa value was 0.29, suggesting a fair to moderate level of agreement among raters (Sim & Wright, 2005).

DISCUSSION

In the literature, the study by Mott et al. (2003) highlighted the significant role of analytic rubrics in enhancing the reliability and validity of writing assessments, emphasizing their effectiveness in evaluating both written and visual narratives. Similarly, Dimopoulos et al. (2013) demonstrated that Learning Analytics-enhanced Rubrics (LAe-R) provide more comprehensive and data-driven assessments by integrating traditional rubric structures with learning analytics, thereby enabling more objective teacher evaluations and deeper insights into student development. Rayon et al. (2014) likewise underscored the importance of enriched rubrics in competency-based assessment by illustrating how student performance data can be systematically analyzed through rubric-based frameworks. In a related vein, Kocakülah (2021) developed a rubric to assess pre-service teachers' problem-solving skills and found that rubric-based assessment supported consistent scoring, accurately captured performance, and positively contributed to academic achievement. Collectively, these studies position rubrics as foundational tools in data-driven and structured assessment practices within education.

Validity and reliability studies on rubrics, alongside research evaluating student performance through rubric-based assessments, consistently emphasize the role of rubrics not only as reliable measurement instruments but also as pedagogical frameworks that guide instruction, clarify learning objectives, and minimize bias in assessment. Rubrics facilitate the integration of assessment with instruction and promote student engagement by supporting self- and peer-assessment practices. Moreover, they are widely regarded as authentic assessment tools that can be embedded in real-world problem-solving tasks (Mertler, 2001; Oakleaf, 2009; Petropoulou, 2011; Brookhart, 2013). In this respect, rubric development research assumes a critical role in the effective integration of artificial intelligence applications into educational contexts.

Within the context of AI-supported education, acquiring prompt-writing skills for effective communication with language-based AI systems and framing these skills within an analytic rubric may facilitate learning processes and address individual learner needs. Although a substantial body of research has explored AI-supported educational practices—such as enhancing motivation, engagement, classroom participation, academic achievement, and decision-making—relatively limited attention has been devoted to the assessment of AI-related competencies through performance-based tools (Arndt, 2023; Khan et al., 2023; Liu et al., 2020; Paek & Kim, 2021; Kumar & Raman, 2022; Winkler & Soellner, 2018; Yılmaz et al., 2021). While previous studies support the use of rubrics for evaluating student performance, they have predominantly focused on conceptual or outcome-oriented dimensions of learning rather than the procedural skills involved in interacting with AI systems. For instance, Kobara et al. (2024) developed analytic rubrics to assess learning outcomes in K–12 AI education; however, their findings revealed low inter-rater reliability, suggesting the need for more context-sensitive and skill-specific assessment criteria. In this regard, the rubric developed in the present study offers a valuable contribution by targeting prompt-writing skills as a measurable and assessable component of AI-supported learning at the K–12 level.

Previous research on assessing learning effectiveness in K–12 artificial intelligence education has largely adopted a knowledge-oriented perspective, focusing on students' conceptual understanding of AI, including dimensions such as understanding of AI, methods of AI use, AI–human relationships, and AI ethics (Kobara et al., 2024). These approaches primarily evaluate students' abilities to explain, relate, and generalize AI-related concepts. In contrast, the rubric developed in the present study addresses a complementary and relatively underexplored dimension of AI education by focusing on students' prompt-writing performance during interactions with language-based AI systems. Rather than assessing what students know about AI, this rubric operationalizes how effectively they can communicate with AI to obtain meaningful and accurate outputs. Accordingly, the emphasis shifts from conceptual AI literacy to an interactional, procedural, and performance-based conception of AI literacy. While previous rubrics evaluate students' explanations of AI usage and ethical considerations, the present rubric captures micro-level writing behaviors that directly shape AI-generated responses, such as clarity and precision in task definition, decomposition of complex requests, construction of appropriate user or role profiles, grammatical accuracy and expressive richness, explicit specification of tone, and systematic output refinement through follow-up prompts. Furthermore, the "Critical Evaluation and Verification of Information" dimension extends beyond general ethical awareness by assessing students' active evaluation of the accuracy and reliability of AI-generated content. By evaluating students' engagement in iterative prompt–response cycles rather than isolated responses, the rubric aligns closely with authentic

classroom practices involving generative AI. Taken together, the rubric does not aim to replace existing AI education assessment tools but rather to complement them by bridging the gap between knowing about AI and using AI productively.

LIMITATIONS

This study has several limitations that should be acknowledged. First, the rubric was validated with a relatively small sample of 32 sixth-grade students from a single private school in Istanbul, which limits the generalizability of the findings to different school types and educational contexts. Second, the study was conducted within a single grade level and instructional setting; therefore, the applicability of the rubric to other grade levels or subject areas remains to be examined. Third, although inter-rater agreement was assessed using two teachers and one researcher, including a larger and more diverse group of raters could provide stronger evidence for scoring consistency. Finally, the rubric was developed based on student interactions with ChatGPT 3.5 within a specific instructional design; thus, its use with other generative AI models or alternative instructional approaches requires further validation.

CONCLUSIONS

In this study, we developed and validated an analytic rubric designed to assess sixth-grade students' prompt-writing skills for effective and accurate communication with ChatGPT. The final rubric consists of five primary dimensions and nine items, and evidence for its validity and reliability was obtained through expert review, exploratory factor analysis, and internal consistency analyses. The factor analysis revealed a robust two-factor structure with satisfactory eigenvalues, explained variance, and factor loadings, while communality values indicated that the extracted factors adequately represented the items. Reliability analyses demonstrated acceptable internal consistency for both factors, and inter-rater agreement, assessed via Fleiss' kappa, indicated a moderate level of consistency among raters. Overall, the findings suggest that the analytic rubric constitutes a valid and reliable tool for evaluating K–12 students' prompt-writing performance and holds promise as a benchmark for AI-integrated educational assessment practices. In addition, the rubric supports the development of students' written communication skills by encouraging clarity, precision, appropriate tone, and purposeful interaction with language-based AI systems.

IMPLICATIONS AND FUTURE RESEARCH

Future research may benefit from employing additional methodological approaches, such as confirmatory factor analysis, reliability generalization, and consistency checking, to further strengthen the evidence for the rubric's validity and reliability (Kamış & Doğan, 2017). Comparative studies examining alternative assessment tools—such as holistic rubrics, checklists, or sequential rating scales—may also provide insights into the most effective strategies for evaluating prompt-writing skills. To enhance inter-rater consistency in classroom implementations, structured teacher training and norming processes are recommended prior to rubric use. Moreover, the analytic rubric developed in this study may be adapted for broader user groups and extended to assess prompt writing for diverse generative AI tasks, including visual or video content generation, thereby supporting future research and practice in this rapidly evolving field.

EFERENCES

Altıntop, M. (2023). Yapay zekâ/akıllı öğrenme teknolojileriyle akademik metin yazma: ChatGPT örneği. *Süleyman Demirel Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 46*, 186–211. https://dergipark.org.tr/tr/pub/sbe/issue/79677/1254533

Amanvermez İncirkuş, F., & Beyreli, L. (2019). A rubric for assessing critical thinking skills through narrative texts. *Journal of Mother Tongue Education, 7*(3), 597–629. https://doi.org/10.16916/aded.553569

Andrade, H. G. (2001). The effects of instructional rubrics on learning to write. *Current Issues in Education, 4*(1), 1–39.

Arndt, H. (2023). AI and education: An investigation into the use of ChatGPT for systems thinking. *arXiv.* https://doi.org/10.48550/arXiv.2307.14206

Arter, J., & McTighe, J. (2001). *Scoring rubrics in the classroom.* Corwin Press.

Aydın, İ. H., & Değirmenci, C. H. (2018). *Yapay zekâ.* Girdap Yayınları.

Bae, J., Kwon, S., & Myeong, S. (2024). Enhancing software code vulnerability detection using GPT-4o and Claude-3.5 Sonnet: A study on prompt engineering techniques. *Electronics, 13*(2657). https://doi.org/10.3390/electronics13132657

Bertalan, M. (2023). Prompt engineering as an important emerging skill for medical professionals: Tutorial. *Journal of Medical Internet Research, 25,* e50638. https://doi.org/10.2196/50638

Bewersdorff, A., Hartmann, C., Hornberger, M., Seßler, K., Bannert, M., Kasneci, E., Kasneci, G., Zhai, X., & Nerdel, C. (2025). Taking the next step with generative artificial intelligence: The transformative role of multimodal large language models in science education. *Learning and Individual Differences, 118,*

102601. https://doi.org/10.1016/j.lindif.2024.102601
Bhatnagar, R., Tanguay, C. L., Sullivan, C., & Many, J. E. (2021). Observation of field practice rubric: Establishing content validity and reliability. *Georgia Educational Researcher, 18*(2), Article 1. https://doi.org/10.20429/ger.2021.180201
Brookhart, S. (2013). *How to create and use rubrics for formative assessment and grading.* ASCD Member Book. ISBN 978-1-4166-1507-1
Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems, 33,* 1877–1901. http://dx.doi.org/10.48550/arXiv.2005.14165
Bıkmaz Bilgen, Ö., & Doğan, N. (2017). Puanlayıcılar arası güvenirlik belirleme tekniklerinin karşılaştırılması [The comparison of interrater reliability estimating techniques]. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 8*(1), 63–78. https://doi.org/10.30831/akukeg.1027601
Büyüköztürk, Ş. (2022). *Sosyal bilimler için veri analizi el kitabı: İstatistik, araştırma deseni, SPSS uygulamaları ve yorum* [Data analysis handbook for social sciences: Statistics, research design, SPSS applications, and interpretation] (29. bs.). Pegem Akademi.
Chakraborty, O., Sahoo, A., Panda, R., & Das, A. (2024). XPL: A cross-model framework for semi-supervised prompt learning in vision-language models. *Transactions on Machine Learning Research, 6.* https://openreview.net/forum?id=oxAZv3QD6M
Chang, C.-Y., Chen, I.-H., & Tang, K.-Y. (2024). *Roles and Research Trends of ChatGPT-based Learning: A Bibliometric Analysis and Systematic Review*. *Educational Technology & Society, 27*(4), 471-486.
Chen, E., Wang, D., Xu, L., Cao, C., Fang, X., & Lin, J. (2024). A systematic review on prompt engineering in large language models for K-12 STEM education. *arXiv.* https://doi.org/10.48550/arXiv.2410.11123
Chiu, T. K. F. (2024). A classification tool to foster self-regulated learning with generative artificial intelligence by applying self-determination theory: A case of ChatGPT. *Educational Technology Research and Development, 72,* 2401–2416. https://doi.org/10.1007/s11423-024-10366-w
Chivers, T. (2023). *Artificial intelligence does not hate you.* Mundi Publishing.
Clark, R. M. (2019). *Intelligence analysis: A target-centric approach* (6th ed.). CQ Press.
Damasio, A. (1999). *Descartes' error.* Varlık Publishing.
Davenport, T. H., & Ronanki, R. (2021). Artificial intelligence for the real world. In *Harvard Business Review Artificial Intelligence Book* (pp. 7–29). Optimist Publishing.
Deniz, K. Z. (Ed.). (2021). *İSTATİSTİKOLAY 2 - Çok değişkenli istatistik.* Nobel Akademi Yayıncılık.
Dimopoulos, I., Petropoulou, O., & Retalis, S. (2013). Assessing students' performance using the learning analytics enriched rubrics. In *LAK '13: Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 195–199). https://doi.org/10.1145/2460296.2460335
Ein-Dor, L., Toledo-Ronen, O., Spector, A., Gretz, S., Dankin, L., Halfon, A., Katz, Y., & Slonim, N. (2024). Conversational prompt engineering. *arXiv.* https://doi.org/10.48550/arXiv.2408.04560
ExcelinEd. (2023). *ChatGPT and education: FAQs for state policymakers.* ExcelinEd Policy Toolkit. https://excelined.org/wp-content/uploads/2023/05/c3_DigitalPolicy_OnePager_ChatGPT-FAQs_2023.pdf
Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). Sage Publications.
Fleiss, J. L., Levin, B., & Paik, M. C. (2003). *Statistical methods for rates and proportions* (3rd ed.). Wiley-Interscience.
Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences.* Basic Books.
Gardner, H. (1999). *Multiple intelligences: Interviews and articles.* BZD Publishing.
Gazzaniga, M. S. (1992). *Nature's mind: The biological roots of thinking, emotions, sexuality, language, and intelligence.* Basic Books Press.
Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning.* MIT Press.
Günbatar, M. S., & Ağgün, N. (2024, September 4–7). Educational applications of artificial intelligence (ChatGPT). In *Proceedings of the 16th National Science and Mathematics Education Congress (UFBMEK 2024)* (pp. 996–997). Van Yüzüncü Yıl University.
Hall, E. W., & Salmon, S. J. (2003). Chocolate chip cookies and rubrics: Helping students understand rubrics in inclusive settings. *Teaching Exceptional Children, 35*(4), 8–13. http://dx.doi.org/10.1177/004005990303500401
Halonen, J. S., & Santrock, J. W. (1996). *Psychology: Contexts of behavior* (2nd ed.). Brown & Benchmark Publishers.
Hill, J., Ford, W. R., & Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. *Computers in Human Behavior, 49,* 245–250. https://doi.org/10.1016/j.chb.2015.02.026
Hoerr, T. (2000). *Becoming a multiple intelligences school.* ASCD.

John, I. (2023). *The art of asking ChatGPT for high-quality answers: A complete guide to prompt engineering techniques.* Nzunda Technologies Limited. ISBN 9781234567890
Jurafsky, D., & Martin, J. H. (2023). *Speech and language processing* (3rd ed.). Prentice Hall.
Khan, B., & Angadi, G. R. (2023). Artificial intelligence integration into school education: A review of Indian and foreign perspectives. *Millennial Asia.* https://doi.org/10.1177/09763996231158229
Knoth, N., Tolzin, A., Janson, A., & Leimeister, J. M. (2024). AI literacy and its implications for prompt engineering strategies. *Computers and Education: Artificial Intelligence, 6,* 100225. https://doi.org/10.1016/j.caeai.2024.100225
Kobara, T., Saito, D., Washizaki, H., & Fukazawa, Y. (2024). Work in progress: Rubrics to assess learning effectiveness of artificial intelligence education for K-12. In *2024 IEEE World Engineering Education Conference (EDUNINE)* (pp. 1–4). https://doi.org/10.1109/EDUNINE60625.2024.10500453
Kocakülah, A. (2022). Development and use of a rubric to assess undergraduates' problem solutions in physics. *Participatory Educational Research, 9*(3), 362–382. http://dx.doi.org/10.17275/per.22.71.9.3
Kula Kartal, S., & Mor Dirlik, E. (2016). The historical development of the concept of validity and the most preferred method for reliability: Cronbach's alpha coefficient. *Abant İzzet Baysal University Journal of Faculty of Education, 16*(4), 1865–1879.
Kumar, V. V. R., & Raman, R. (2022). Student perceptions on artificial intelligence (AI) in higher education. In *2022 IEEE Integrated STEM Education Conference (ISEC)* (pp. 1–7). IEEE. https://doi.org/10.1109/ISEC51903.2022.9745086
Kurzweil, R. (2005). *The singularity is near: When humans transcend biology.* Viking Press.
Kutlucan, E., & Seferoğlu, S. S. (2024). The use of artificial intelligence in education: A KEFE and PEST analysis of ChatGPT. *Turkish Journal of Educational Sciences, 22*(2), 1059–1083. https://doi.org/10.37217/tebd.1368821
Köksal, A. (2007). Üstün zekalı çocuklarda duygusal zekâyı geliştirmeye dönük program geliştirme çalışması [A program development study for developing emotional intelligence in gifted children] (Doktora tezi, İstanbul Üniversitesi, Eğitim Bilimleri Enstitüsü).
Köse, U. (2022). *Yapay zeka felsefesi.* Doğu Kitapevi.
Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology, 28*(4), 563–575. https://psycnet.apa.org/doi/10.1111/j.1744-6570.1975.tb01393.x
Li, H., & Klabjan, D. (2024). Reverse prompt engineering. *arXiv.* https://doi.org/10.48550/arXiv.2411.06729
Li, J., Jangamreddy, N. K., Hisamoto, R., Bhansali, R., Dyda, A., Zaphir, L., & Glencross, M. (2024). AI-assisted marking: Functionality and limitations of ChatGPT in written assessment evaluation. *Australasian Journal of Educational Technology, 40*(4), 56–72. https://doi.org/10.14742/ajet.9463
Lindley, J., & Whitham, R. (2024). From prompt engineering to prompt craft. *arXiv.* https://doi.org/10.48550/arXiv.2411.13422
Liu, J., Chang, H., Forrest, J. Y.-L., & Yang, B. (2020). Influence of artificial intelligence on technological innovation: Evidence from the panel data of China's manufacturing sectors. *Technological Forecasting and Social Change, 158,* 120142. https://doi.org/10.1016/j.techfore.2020.120142
Marr, B. (2021). *The artificial intelligence revolution.* Optimist Publishing Group.
McCarthy, J. (2007, November 12). What is artificial intelligence? *Stanford University.* http://www-formal.stanford.edu/jmc/
Melanson, J., & Maman, B. (2024). *ChatGPT complete guide: Learn Midjourney, DALL-E 2 & more.* Self-published.
Mertler, C. A. (2001). Designing scoring rubrics for your classroom. *Practical Assessment, Research & Evaluation, 7*(25). https://doi.org/10.7275/gcy8-0w24
Mesko, B. (2023). The impact of multimodal large language models on health care's future. *Journal of Medical Internet Research, 25,* e52865. https://doi.org/10.2196/52865
Mhlanga, D. (2023). Open AI in education, the responsible and ethical use of ChatGPT towards lifelong learning. *SSRN Electronic Journal.* https://dx.doi.org/10.2139/ssrn.4354422
Ministry of National Education [Millî Eğitim Bakanlığı]. (2024). Regulation on measurement and evaluation of the Ministry of National Education: Part one, preliminary provisions [Millî Eğitim Bakanlığı ölçme ve değerlendirme yönetmeliği: Birinci bölüm, genel hükümler]. Ankara, Turkey.
Mitchell, T. (1997). *Machine learning.* McGraw-Hill.
Mollick, E., & Mollick, L. (2023). *The AI classroom: The ultimate guide to artificial intelligence in education.* Jossey-Bass.
Moskal, B. M., & Leydens, J. A. (2000). Scoring rubric development: Validity and reliability. *Practical Assessment, Research & Evaluation, 7*(10).
Mott, M. S., Etsler, C., & Drumgold, D. (2003). Applying an analytic writing rubric to children's hypermedia "narratives." *Early Childhood Research & Practice, 5*(1). https://doi.org/10.1177/2158244012445584
Myers, D. G. (1998). *Psychology* (5th ed.). Worth Publishers.

Oakleaf, M. (2009). Using rubrics to assess information literacy: An examination of methodology and interrater reliability. *Journal of the American Society for Information Science and Technology, 60*(5), 969–983. https://doi.org/10.1002/asi.21030
OpenAI. (n.d.). Prompt engineering. https://platform.openai.com/
Oz, H. (2023). *ChatGPT & artificial intelligence: Prompt engineering from zero to mastery* [Online course]. Udemy. https://www.udemy.com/course/chatgpt-ai-verimliligi-artrmak-icin-prompt-muhendisligi/
Paek, S., & Kim, N. (2021). Analysis of worldwide research trends on the impact of artificial intelligence in education. *Sustainability, 13*(14), 7941. https://doi.org/10.3390/su13147941
Petropoulou, O., Vassilikopoulou, M., & Retalis, S. (2011). Enriched assessment rubrics: A new medium for enabling teachers to easily assess students' performance when participating in complex interactive learning scenarios. *Operational Research, 11*(2), 171–186. http://dx.doi.org/10.1007/s12351-009-0047-5
Pranab, S., Ayush, K. S., Sriparna, S., Vinija, J., Samrat, M., & Aman, C. (2024). A systematic survey of prompt engineering in large language models: Techniques and applications. *arXiv.* https://doi.org/10.48550/arXiv.2402.07927
Prompting Guide. (2025). *Prompting guide.* OpenAI. https://cookbook.openai.com/
Qureshi, B. (2023). Exploring the use of ChatGPT as a tool for learning and assessment in undergraduate computer science curriculum: Opportunities and challenges. *arXiv.* https://doi.org/10.48550/arXiv.2304.11214
Rau, W. A. (2001). The relationship of emotional intelligence test scores to job performance evaluation scores in the management group of a health care organization (Unpublished doctoral dissertation). Medical University of South Carolina.
Rayon, A., Guenaga, M., & Núñez, A. (2014). Supporting competency assessment through a learning analytics approach using enriched rubrics. In *TEEM '14: Proceedings of the Second International Conference on Technological Ecosystems for Enhancing Multiculturality* (pp. 291–298).
Reeves, A. R. (2011). *Where great teaching begins: Planning for student thinking and learning.* ASCD Press.
Riggio, R. E., Murphy, S. E., & Pirozzolo, F. J. (Eds.). (2002). *Multiple intelligences and leadership.* Lawrence Erlbaum Associates.
Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson.
Sander, S., Michael, I., Nishant, B., Konstantine, K., Amanda, L., Chenglei, S., Yinheng, L., Aayush, G., HyoJung, H., Sevien, S., et al. (2024). The prompt report: A systematic survey of prompting techniques. *arXiv.* https://doi.org/10.48550/arXiv.2406.06608
Say, C. (2018). *Artificial intelligence in 50 questions.* Bilim ve Gelecek Library.
Schoepp, K., Danaher, M., & Ater Kranov, A. (2018). An effective rubric norming process. *Practical Assessment, Research, and Evaluation, 23*(1), 11. https://doi.org/10.7275/z3gm-fp34
Schulhoff, S., Ilie, M., Balepur, N., Kahadze, K., Liu, A., Si, C., ... & Resnik, P. (2024). The prompt report: A systematic survey of prompting techniques. *arXiv.* https://doi.org/10.48550/arXiv.2406.06608
Shariff, A. F., B. S., A., Radhika, M., & S, S. (2020). Blind assistance using artificial intelligence. *International Journal of Engineering Applied Sciences and Technology, 5*(3), 396–401. http://dx.doi.org/10.33564/IJEAST.2020.v05i03.063
Shidiq, M. (2023). The use of artificial intelligence-based ChatGPT and its challenges for the world of education: From the viewpoint of the development of creative writing skills. In *Proceedings of the International Conference of Education, Society and Humanity, 1*(1).
Sim, J., & Wright, C. C. (2005). The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy, 85*(3), 257–268. [PMID: 15733050]
Solso, R. L. (1995). *Cognitive psychology* (4th ed.). Allyn and Bacon.
Spasić, A. J., & Janković, D. S. (2023). Using ChatGPT standard prompt engineering techniques in lesson preparation: Role, instructions, and seed-word prompts. In *Proceedings of the International Conference on Emerging eLearning Technologies and Applications (ICEST).* IEEE. https://doi.org/10.1109/icest58410.2023.10187269
Sternberg, R. J. (1985). *Human abilities: An information processing approach.* W. H. Freeman & Co Press.
Sternberg, R. J. (2004). North American approaches to intelligence. In R. J. Sternberg (Ed.), *International handbook of intelligence* (pp. 411–444). Cambridge University Press.
Sternberg, R. J. (2005). *Intelligence and intelligence testing.* Cambridge University Press.
Sutarso, P. (1998). Gender differences on the emotional intelligence inventory (Unpublished doctoral dissertation). University of Alabama.
Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press.
Taggart. (1999). *Rubrics: A handbook for construction and use.* R&L Education.
Thorndike, E. L. (1920). Intelligence and its uses. *Harper's Magazine, 140,* 227–235.
Tom, B. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G.,

Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). Language models are few-shot learners. *arXiv.* https://arxiv.org/abs/2005.14165

Turing, A. M. (1950). Computing machinery and intelligence. *Mind, 59*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

Vairamani, A. D., & Nayyar, A. (2024). *Prompt engineering: Empowering communication.* CRC Press.

Vatansever, A. N. (2024). Üniversite öğrencilerinin yapay zekâ kavramına ilişkin metaforları ve görüşleri üzerine karşılaştırmalı nitel bir araştırma [A comparative qualitative study on university students' metaphors and opinions regarding the concept of artificial intelligence] (Yüksek Lisans tezi, Marmara Üniversitesi, Eğitim Bilimleri Enstitüsü). Council of Higher Education Thesis Center.

Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: The kappa statistic. *Family Medicine, 37*(5), 360–363. PMID: 15883903

Wechsler, D. (1943). Non-intellective factors in general intelligence. *The Journal of Abnormal and Social Psychology, 38*(1), 101–103. https://doi.org/10.1037/h0060613

Winkler, R., & Soellner, M. (2018). Unleashing the potential of chatbots in education: A state-of-the-art analysis. *Academy of Management Proceedings, 2018*(1), 15903. https://doi.org/10.5465/AMBPP.2018.15903

Woodcock, R. W. (1990). Theoretical foundations of the WJ-R measures of cognitive ability. *Journal of Psychoeducational Assessment, 8*(3). https://doi.org/10.1177/073428299000800303

Xu, W., & Ouyang, F. (2022). The application of AI technologies in STEM education: A systematic review from 2011 to 2021. *International Journal of STEM Education, 9*(59). https://doi.org/10.1186/s40594-022-00377-5

Yılmaz, Y., Uzelli Yılmaz, D., Yıldırım, D., Akın Korhan, E., & Özer Kaya, D. (2021). Opinions of faculty of health sciences students on artificial intelligence and its use in healthcare [Yapay zekâ ve sağlıkta yapay zekânın kullanımına yönelik Sağlık Bilimleri Fakültesi öğrencilerinin görüşleri]. *Süleyman Demirel University Journal of Health Sciences, 12*(3), 297–308. https://doi.org/10.22312/sdusbed.950372

Yılmaz, A. (2021). *Yapay zeka* [E-book]. KODLAB Yayınları.

Zamfirescu-Pereira, J. D., Wei, H., Xiao, A., Gu, K., Jung, G., Lee, M. G., Hartmann, B., & Yang, Q. (2023). Herding AI cats: Lessons from designing a chatbot by prompting GPT-3. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference (DIS '23)* (pp. 2206–2220). Association for Computing Machinery.

Appendix A

AI prompt writing rubric

Dimensions	Excellent (4)	Proficient (3)	Partially Proficient (2)	Not proficient (1)	SCORE
Clarity and Precision in Task Definition	Always articulates their requests clearly when defining the task to a language-based AI.	Usually articulates their requests clearly when defining the task to a language-based AI.	Sometimes articulates their requests clearly when defining the task to a language-based AI.	Never articulates their requests clearly when defining the task to a language-based AI.	
	Always breaks down their requests into small, manageable components as concise sentences when defining the task to a language-based AI.	Usually breaks down their requests into small, manageable components as concise sentences when defining the task to a language-based AI.	Sometimes breaks down their requests into small, manageable components as concise sentences when defining the task to a language-based AI.	Enter their requests in a single, long, and hard-to-understand sentence when defining the task.	
Profile Creation	Always create a level-appropriate profile to complete the task and add details that elaborate on the task.	Usually create a level-appropriate profile to complete the task and add some details that elaborate on the task.	Does not create a profile when defining the task, but the task itself is clear and comprehensible.	Does not create a profile when defining the task and fails to express the task clearly and comprehensibly.	
Grammar Usage and Expression	Always provides accurate commands when defining the task for the language-based AI.	Usually provides accurate commands when defining the task for the language-based AI.	Sometimes provides accurate commands when defining the task for the language-based AI.	Provides commands using only a single word or phrase without any action-oriented element when defining the task for the language-based AI.	
	Enhances the effectiveness of the prompt by always using adjectives, conjunctions, and adverbs correctly.	Enhances the effectiveness of the prompt by usually using adjectives, conjunctions, and adverbs correctly.	Sometimes uses adjectives, conjunctions, and adverbs correctly. The request clarity is poor.	Never uses adjectives, conjunctions, and adverbs when defining the task.	
	Always adopts a tone consistent with the defined task. Ensures a more accurate output by clearly stating this tone in the prompt (e.g., casual-informal tone or formal-academic tone).	Usually adopts a tone consistent with the defined task.	Sometimes adopts a tone consistent with the defined task.	Never adopts a tone when defining the task.	
Output Refinement and Attention to Details	Always evaluates the generated output.	Usually evaluates the generated output.	Sometimes evaluates the generated output.	Never evaluates the generated output.	
	Always writes follow-up prompts with refined details to complete or improve nuanced aspects of the output.	Usually writes follow-up prompts with refined details to complete or improve nuanced aspects of the output.	Sometimes writes follow-up prompts with refined details to complete or improve nuanced aspects of the output.	Never writes follow-up prompts with refined details. Uses the output as it is without any refinement.	

Critical Evaluation and Verification of Information	Always critically evaluates the accuracy and reliability of the information provided by the language-based AI.	Usually critically evaluates the accuracy and reliability of the information provided by the language-based AI.	Sometimes critically evaluates the accuracy and reliability of the information provided by the language-based AI.	Never critically evaluates the accuracy and reliability of the information provided by the language-based AI.	

Authentic and Transformational Leadership in Educational Management: A Comparative Review and Synthesis

Taleh Mirzayev
Azerbaijan State Pedagogical University, Department of Public Relations and Marketing, Baku, Azerbaijan
taleh.mirzayev@adpu.edu.az

Abstract
This study examines authentic and transformational leadership in educational settings by integrating systematic literature findings with large-scale trend analyses from Web of Science, Scopus, and Scopus AI. Drawing on conceptual foundations, empirical evidence, and AI-supported semantic mapping, the study analyzes the core dimensions, mechanisms, teacher-related outcomes, and cultural contexts associated with both leadership styles. Results indicate that authentic leadership primarily strengthens teachers' autonomy, well-being, trust, and ethical climate through mechanisms such as relational transparency, emotional support, and organizational justice. Transformational leadership, meanwhile, exerts a stronger influence on innovation, collective efficacy, job satisfaction, and organizational commitment, driven by inspirational motivation, intellectual stimulation, and shared vision.
Database trends show that research on transformational leadership is substantially more extensive and historically rooted, whereas authentic leadership has emerged more recently and remains less represented, particularly in longitudinal designs. Scopus AI analysis reveals three dominant thematic clusters: authentic leadership, transformational leadership, and comparative or integrative studies. Both leadership styles exhibit significant conceptual overlap, yet they differ in their psychological pathways and contextual sensitivity. Cross-cultural variability highlights the moderating role of educational governance, collectivism, and policy structures.
The study identifies key gaps, including the scarcity of longitudinal and hybrid models, limited cross-cultural comparative work, and measurement challenges due to construct overlap. It concludes that integrating authentic and transformational leadership may offer a more comprehensive framework for fostering sustainable innovation, teacher well-being, and ethical school cultures.
Keywords: *Authentic leadership, Transformational leadership, Educational leadership, Scopus AI analysis, Comparative leadership studies.*

Introduction

Leadership research historically evolved from trait theories, behavioral approaches, and contingency models; however, since the 1980s, ethical scandals, increasing demands for organizational change, and the complexity of human interaction have shifted leadership studies from task-centered frameworks toward relational, ethical, and transformational paradigms (Avolio & Gardner, 2005; Cemaloğlu & Özdemir, 2019). Within this modern landscape, authentic leadership and transformational leadership have become central models positioned at the intersection of positive organizational behavior, ethics, and organizational psychology (Walumbwa et al., 2008; Çevik, 2024).
Both approaches conceptualize leadership as more than administrative task completion; leaders are framed as psychological, moral, and relational agents capable of shaping followers' identities, motivations, and organizational meaning systems (Anderson, 2017; Kareem et al., 2023). This is particularly salient in educational institutions, where human development is the core organizational mission.

Aim

The purpose of this study is to systematically examine the theoretical foundations, mechanisms, and educational implications of authentic leadership (AL) and transformational leadership (TL), and to compare their effects on teacher outcomes, organizational culture, and institutional innovation. Drawing on empirical evidence and conceptual developments from the past two decades, the study aims to synthesize findings from Web of Science (WoS) and Scopus-indexed publications, including trend analyses generated through Scopus AI. A further aim is to identify conceptual overlap, mediating mechanisms, contextual moderators, and methodological gaps-particularly the scarcity of longitudinal and integrated leadership models in educational research. By doing so, the study contributes to building a comprehensive and analytically differentiated understanding of how AL and TL shape educational environments and to illuminate future directions for research and practice.

Significance

The significance of this study lies in the growing recognition that leadership is among the most influential determinants of teacher well-being, professional growth, school climate, and student achievement (Anderson, 2017; Kareem et al., 2023). Although authentic and transformational leadership have emerged as two of the most prominent paradigms in contemporary educational leadership, the existing literature reveals substantial conceptual overlap and an uneven distribution of research emphasis (Çevik, 2024; Xie, Ahmad, & Lu, 2024a).

Transformational leadership is far more extensively studied, with publication volumes in WoS and Scopus up to seven times higher than those of authentic leadership. This imbalance reflects both historical precedence (Bass, 1985; Burns, 1978) and methodological convenience, as TL outcomes are more readily measurable using validated instruments such as the MLQ-5X. In contrast, authentic leadership-despite its importance for psychological well-being, ethical climate, relational trust, and transparency-remains less empirically explored and is often investigated through cross-sectional designs (Walumbwa et al., 2008; Hsu et al., 2024).

This study is significant because it brings together fragmented scholarly insights through a comparative, mechanism-based, and context-sensitive synthesis, revealing:

- how AL and TL differentially influence teacher autonomy, innovation, organizational commitment, and psychological capital;
- the mediating roles of trust, self-efficacy, collective efficacy, and organizational culture;
- the moderating effects of cultural norms, collectivism, and policy structures;
- the need for longitudinal and hybrid leadership models to reflect real-time school dynamics.

By integrating database trends, empirical findings, and conceptual frameworks, this study provides a more holistic understanding of leadership in education and identifies avenues for more precise, culturally responsive, and methodologically rigorous future research.

Method

This study adopts an integrated multi-stage research design combining systematic review procedures, database-driven trend analytics, and concept-level synthesis informed by Scopus AI. The methodology was updated to incorporate the analytical categories reflected in Table 8, including leadership dimensions, teacher impact, mediators, measurement tools, cultural contexts, and longitudinal research gaps.

1. Systematic Literature Review

A structured review of peer-reviewed research on authentic (AL) and transformational leadership (TL) in educational settings was conducted. The review included:

- conceptual and empirical studies on AL and TL,
- comparative and integrated leadership research,
- studies examining mechanisms such as self-efficacy, trust, and organizational culture,
- research addressing teacher outcomes (e.g., autonomy, well-being, innovation, job satisfaction).

The dataset included 13 core open-access studies plus additional seminal works across leadership theory.

2. Web of Science (WoS) and Scopus Trend Analysis

To capture broader scholarly patterns, targeted searches were conducted in WoS and Scopus using:

- *"authentic leadership" AND education*
- *"transformational leadership" AND education"*

Inclusion criteria:

- peer-reviewed journal articles
- English language
- social sciences, education, management, psychology, and multidisciplinary fields
- publication years 1992–2026 (database-specific)

The extracted data were organized according to:

- annual publication distribution
- disciplinary distribution
- country contributions
- volume comparison of AL vs. TL studies

These results informed the comparative portions of Table 8.

3. Scopus AI–Supported Meta-Analytic Concept Mapping
A Scopus AI–driven semantic analysis was performed to identify:

- thematic clusters around AL, TL, and comparative/integrative models
- leadership dimensions (e.g., self-awareness, idealized influence)
- organizational and psychological mediators (e.g., autonomy, collective efficacy, emotional intelligence)
- measurement tools (ALQ, MLQ-5X, GTLS)
- cultural contextual moderators (e.g., collectivism, policy systems)
- gaps in longitudinal and integrated research

This analysis directly shaped the structure of Table 7 and Table 8.

4. Comparative Synthesis Based on Table 8 Categories
Using the dimensions of Table 8, the integrative analysis compared AL and TL along the following domains:

- Core dimensions (AL: self-awareness, transparency, moral perspective; TL: idealized influence, inspiration, intellectual stimulation)
- Impact on teachers (AL → well-being, autonomy, ethical climate; TL → innovation, self-efficacy, commitment)
- Mediators (e.g., trust, psychological safety, collective efficacy)
- Measurement instruments (ALQ vs. MLQ-5X/GTLS)
- Cultural contexts (collectivism, centralization, policy frameworks)
- Longitudinal/integrated research gaps

These dimensions were used to build cross-model comparisons and identify fine-grained construct overlap ($\rho \approx .72$) and differentiation needs.

5. Narrative Integration and Research Gap Identification
Given the heterogeneity of the sources (empirical, conceptual, AI-derived), a narrative synthesis was conducted to:

- integrate AL and TL contributions to teacher outcomes,
- analyze mediating organizational cultures,
- explore emotional intelligence as a cross-cutting moderator,
- compare cross-cultural findings across education systems,
- highlight methodological limitations (dominance of cross-sectional designs),
- identify research needs for hybrid models and longitudinal designs.

Ethical approval was not required, as the study used secondary data.

Theoretical Foundations
Theoretical Foundations of Authentic Leadership
Authenticity emphasizes congruence between a leader's values, identity, and actions. The conceptual roots of authentic leadership stem from positive psychological development (Avolio & Luthans, 2003), ethical leadership, and self-regulation theories (Kernis, 2003). Walumbwa et al.'s (2008) influential four-dimensional model-self-awareness, relational transparency, balanced information processing, and internalized moral perspective-frames authentic leadership as a value-driven and trust-oriented approach.

Authentic leaders foster alignment between inner values and outward behaviors, thereby enhancing perceptions of integrity, credibility, and psychological safety (Khanyile & De Bruin, 2022). Theoretically, authentic leadership intersects with positive organizational behavior (Luthans & Youssef, 2004), ethical leadership, and social learning theory (Bandura, 1997), all of which emphasize moral modeling and relational consistency.

The core assumption is that when leaders act in accordance with their genuine selves, followers internalize both the leader and the organizational mission more readily, strengthening psychological well-being, moral agency, and organizational commitment (Xie, Ahmad, & Lu, 2024a).

Empirical Findings in Educational Contexts
Recent educational studies provide robust support for the relational and psychological effects of authentic leadership.

- Feng (2016) found, in a study of 1,429 teachers, that all four dimensions of authentic leadership significantly enhance teachers' psychological capital, especially hope, resilience, and optimism.

- Hsu et al. (2024) reported that self-awareness, relational transparency, and balanced processing substantially increase teachers' creativity and job performance.
- Systematic reviews confirm its positive association with trust, ethical climate, participatory decision-making, and reduced burnout (Çevik, 2024; Khanyile & De Bruin, 2022).

However, authentic leadership shows weaker effects on innovation and rapid change, suggesting a contextual limitation in dynamic reform environments.

Theoretical Foundations of Transformational Leadership

Transformational leadership originates from Burns' (1978) distinction between transactional and transformational leadership and is deeply informed by social change theory. Bass (1985) operationalized the model through four dimensions: (1) Idealized influence, (2) Inspirational motivation, (3) Intellectual stimulation, (4) Individualized consideration.

Bass and Avolio (1994) later integrated these in the Full Range Leadership Model, positioning transformational leadership alongside transactional and laissez-faire styles within a broader system.

Early contributors such as Downton (1973) and House (1977) emphasized charisma, value-driven influence, and visionary orientation, situating transformational leadership as a strategic, motivational, and future-oriented paradigm. This model has been widely supported across organizational contexts, particularly due to its capacity to mobilize followers toward collective goals, identity transformation, and innovation (Gumusluoglu & Ilsev, 2009; Williams, 2016).

Empirical Findings in Educational Contexts

Transformational leadership is among the most widely researched models in education, consistently linked to:

- higher teacher motivation and organizational commitment (Kareem et al., 2023),
- enhanced school performance (Ramos, 2025),
- stronger institutional innovation (Elrehail et al., 2018),
- organizational learning and knowledge sharing (Anderson, 2017).

A notable finding from Elrehail et al. (2018) is that transformational leadership strongly predicts both process and product innovation, especially when knowledge-sharing norms are high. The same study reported that authentic leadership showed no significant relationship with innovation, highlighting a key divergence between the two models.

Comparison of Authentic and Transformational Leadership

1. Focus and Purpose

- Authentic leadership prioritizes moral grounding, self-awareness, ethical consistency, and relational transparency (Walumbwa et al., 2008).
- Transformational leadership centers on organizational vision, collective motivation, and change mobilization (Bass, 1985).

2. Theoretical Foundations

- Authentic leadership: positive psychology, ethical leadership, self-regulation, values-based leadership (Avolio & Luthans, 2003).
- Transformational leadership: social change theory, charismatic leadership, and the Full Range Leadership Model (Burns, 1978; Bass & Avolio, 1994).

3. Psychological and Organizational Outcomes

- Authentic leadership → trust, ethical climate, psychological capital, well-being, relational cohesion (Khanyile & De Bruin, 2022; Xie et al., 2024b).
- Transformational leadership → motivation, innovation, performance, organizational commitment, pedagogical change (Anderson, 2017; Ramos, 2025).

4. Innovation and Change

- Transformational leadership is strongly innovation-oriented (Gumusluoglu & Ilsev, 2009; Elrehail et al., 2018).
- Authentic leadership has limited or context-dependent effects on innovation.

5. Educational Context Implications

- Authentic leadership supports ethical climate, emotional well-being, and trust among teachers.
- Transformational leadership drives structural change, digitalization, school performance, and reform implementation.

Toward an Integrated Authentic–Transformational Leadership Synthesis

Contemporary literature increasingly emphasizes that the two models should not be viewed as mutually exclusive but as complementary (Çevik, 2024; Anderson, 2017).

Authentic leadership provides the ethical, relational, and psychological foundation necessary for sustainable leadership practices. Transformational leadership builds on this foundation to mobilize change, foster innovation, and produce measurable organizational outcomes.

Thus, the most effective educational leadership model is a hybrid authentic–transformational orientation, which combines:

- authenticity → trust, ethics, transparency, psychological safety;
- transformational capacity → vision, innovation, collective motivation, organizational learning.

This synthesis aligns with the dual nature of contemporary educational institutions, which require both ethical-relational stability and strategic-innovative dynamism.

Trends in Current Research on Authentic and Transformational Leadership in Education: A Web of Science and Scopus–Based Review

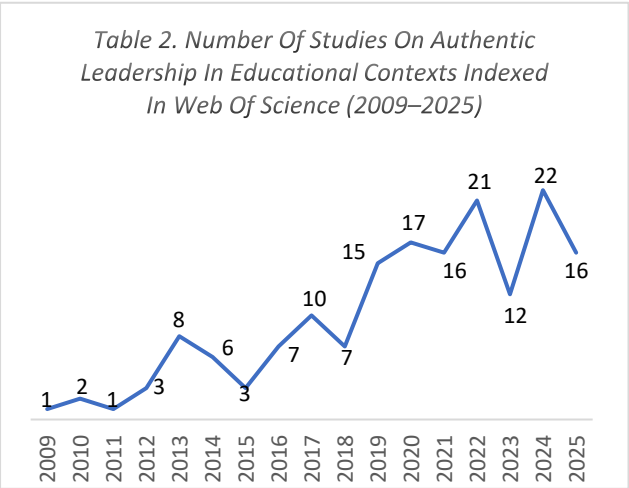
In addition to the primary literature reviewed above, a complementary database analysis was conducted using Web of Science (WoS) and Scopus to identify current trends in educational research on authentic and transformational leadership. Searches were performed using the keywords “authentic leadership” AND education and “transformational leadership” AND education, limited to peer-reviewed journal articles written in English within the social sciences, humanities, management sciences, and multidisciplinary fields.

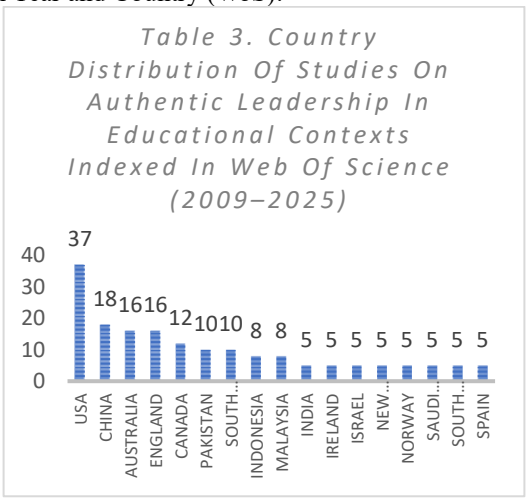
Findings on Authentic Leadership in Educational Research

Based on the predefined inclusion criteria, the Web of Science database contains 167 peer-reviewed journal articles addressing authentic leadership in educational contexts. These studies are predominantly situated within Education (84 articles, 50.29%) and Management (50 articles, 29.94%). The earliest article indexed in WoS on this topic dates back to 2009.

Table 1. Fields of Publications on Authentic Leadership in Education in WoS (2009–2025)

Table 2–3. Distribution by Publication Year and Country (WoS).



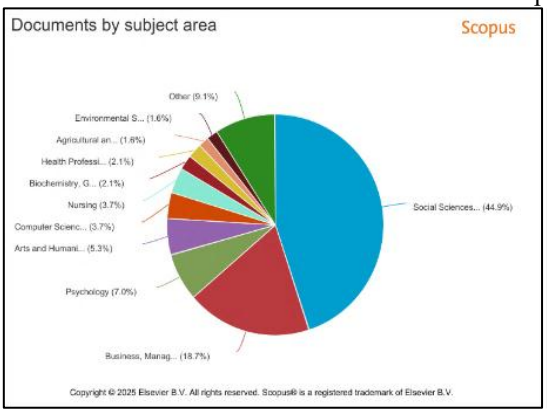


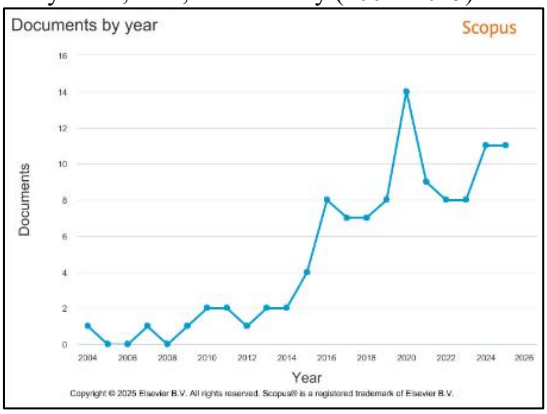
The tables indicate that the scholarly discourse around authentic leadership in education has emerged prominently over the past 15–20 years, with a significant growth in the last decade. Specifically, 119 of the 167 articles (71.26%) were published within the last seven years, demonstrating increasing academic attention.

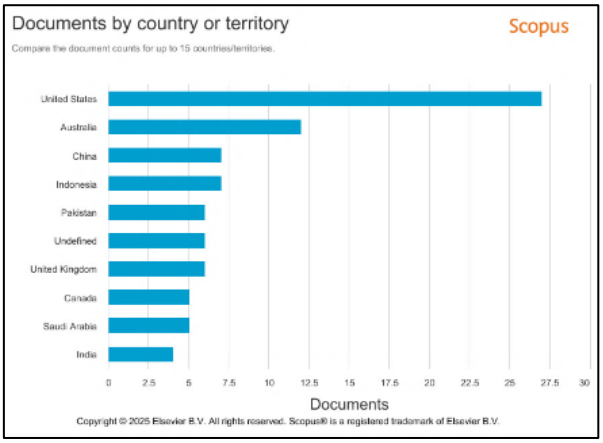
In terms of geographic distribution, the majority of publications originate from the United States, China, Australia, the United Kingdom, and Canada, with Türkiye ranked 12th with three publications.

Scopus Trends on Authentic Leadership in Education

Table 4. Distribution of Publications in Scopus by Field, Year, and Country (2004–2025)







As shown in Table 4, similar tendencies exist in Scopus. A total of 107 peer-reviewed English-language articles are indexed, spanning a broader time range (2004–2025). Most studies fall within:

— Social Sciences (44.9%) and Management and Business (18.7%), with the remaining percentage distributed across psychology, education, and cross-disciplinary fields.

Importantly, 91 of the 107 articles (85.05%) have been published within the last ten years, indicating accelerated interest. As in WoS, the United States leads the field, followed by Australia, China, Indonesia, and Pakistan. Notably, only one article in Scopus originates from a researcher affiliated with Türkiye.

Trends in Transformational Leadership Research in Education

To provide a comparative understanding, a parallel analysis was conducted for transformational leadership.

Table 5. Web of Science Publications on Transformational Leadership in Education (1992–2025)

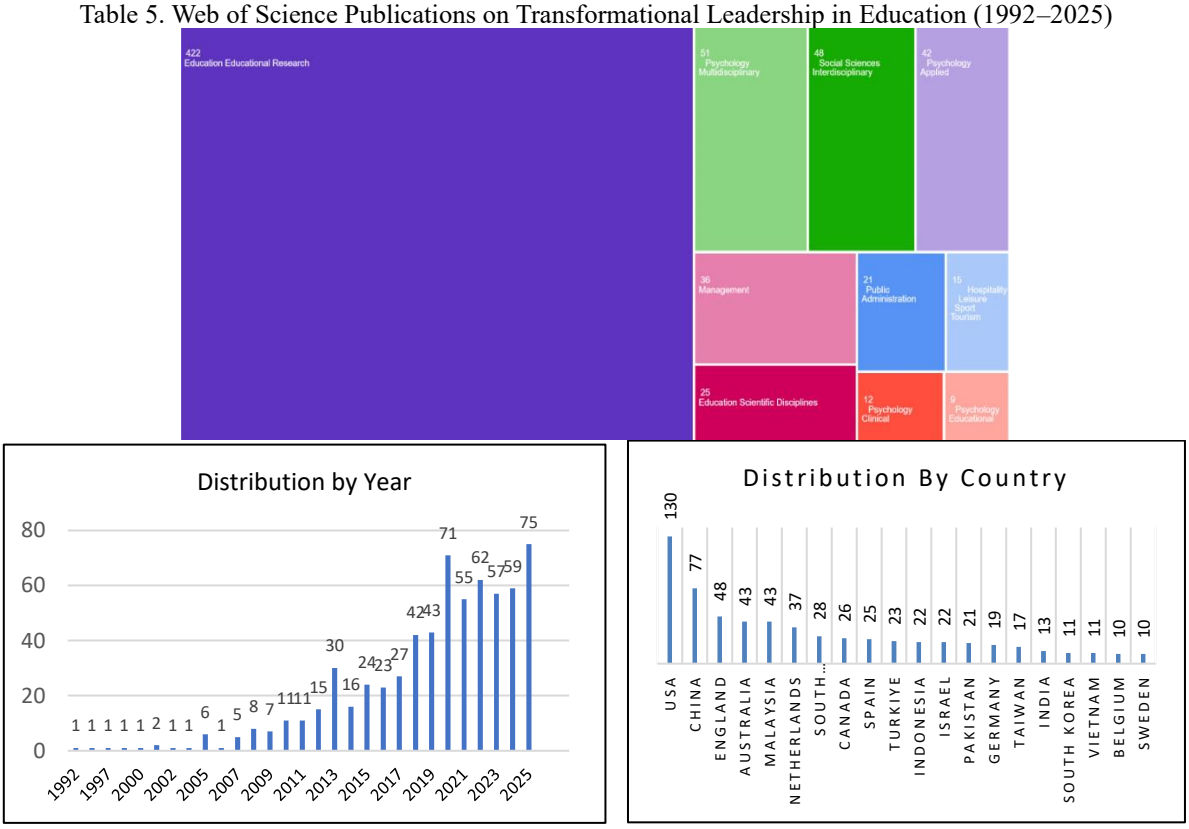
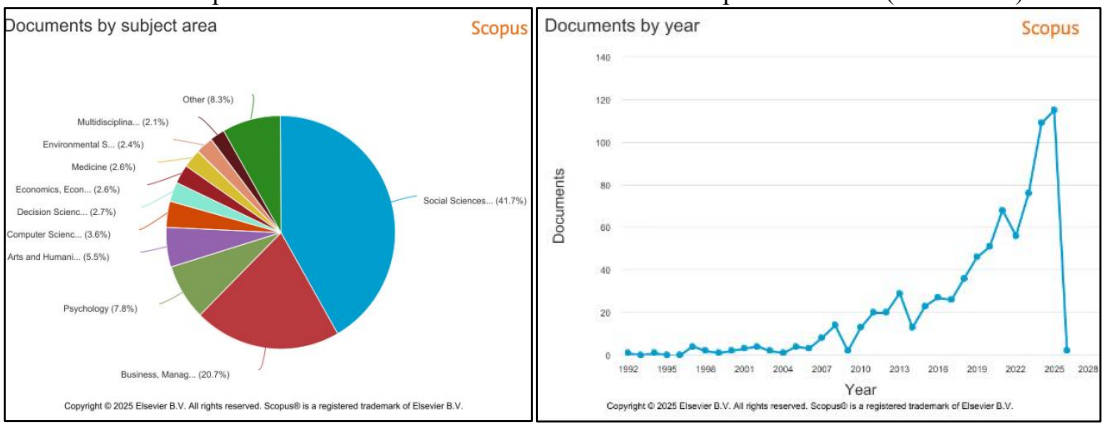
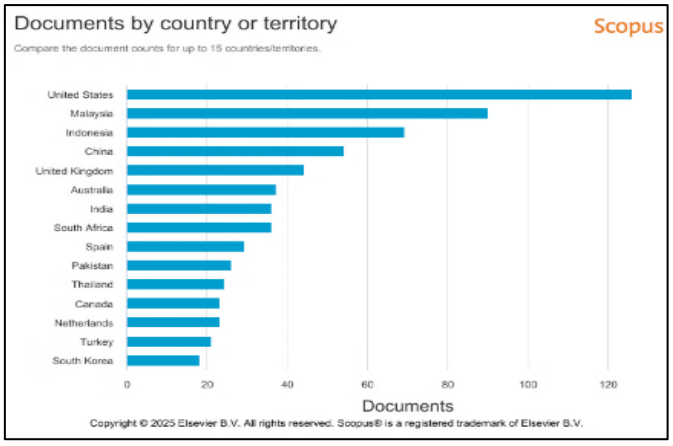


Table 6. Scopus Publications on Transformational Leadership in Education (1992–2026)





The findings reveal that transformational leadership has generated significantly more research attention compared to authentic leadership:

- 657 articles in Web of Science (≈4 times more than authentic leadership),
- 784 articles in Scopus (≈7 times more).

The earliest studies appear in 1992, indicating a much longer research tradition. Publication frequency increases sharply beginning in 2013, with the highest concentration of studies emerging in the last five years.
This trend suggests that transformational leadership is more extensively theorized and empirically tested in educational settings.

Reasons for Greater Research Volume
Two primary explanations account for this difference:

1. Transformational leadership predates authentic leadership, allowing for a longer period of theoretical development and empirical exploration.
2. Transformational leadership outcomes are more easily measurable, making it more suitable for quantitative research designs. Authentic leadership, however, typically requires more nuanced, complex, and labor-intensive qualitative methodologies, limiting its volume of empirical work.

Field Distribution
The disciplinary distribution mirrors that of authentic leadership yet with higher density:

- In WoS, 64.23% of transformational leadership studies fall within education.
- In Scopus, 41.7% are in social sciences, and 20.7% are in management.

Geographical Trends
While the United States remains the leading contributor, subsequent positions are occupied by Australia, Indonesia, Pakistan, and Türkiye, among others. Türkiye demonstrates notable scholarly engagement:

- 10th in WoS with 23 publications,
- 14th in Scopus with 21 publications.

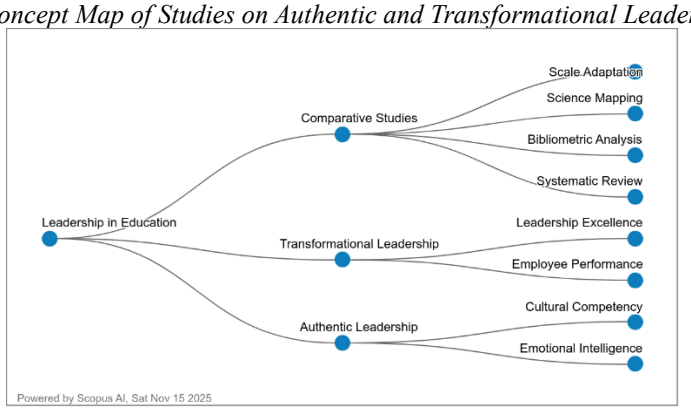
This reflects both the increasing involvement of researchers in Eastern contexts and an expanding interest in the cultural relevance of leadership models.

An additional observation based on Scopus's AI topic analysis indicates that recent influential contributors (e.g., Mingyu Hou, Yi Zhao, Jamilah Binti Ahmad) predominantly examine the impact of transformational leadership on school-level outcomes, further reinforcing its contemporary relevance in global educational research agendas.

Scopus AI–Supported Field Analysis
Leadership in education is a critical determinant of teacher development, institutional culture, and student outcomes. Among various leadership paradigms, authentic and transformational leadership have received substantial attention due to their potential to foster innovation, well-being, and professional engagement among educators. The present analysis aims to summarize current research trends in Scopus pertaining to these leadership styles, their dimensions, mediating mechanisms, and comparative effects, and to identify key gaps and future research directions in the field of educational leadership.

The analysis is based on insights generated through Scopus AI, which provides an advanced meta-analytic synthesis of studies indexed in the Scopus database. Below is the concept map model illustrating the thematic structure of authentic and transformational leadership research within Scopus.

Table 7. Concept Map of Studies on Authentic and Transformational Leadership Indexed in Scopus



As illustrated in the concept map, research on these leadership styles in Scopus converges around three primary thematic areas:

1. Transformational leadership,
2. Authentic leadership,
3. Comparative or integrative studies.

The following section provides a closer summary of the research emerging within these themes.

Dimensions and Mechanisms of Authentic and Transformational Leadership in Education

Transformational leadership is grounded in cultivating a shared vision, supporting organizational change, and encouraging educators to transcend conventional boundaries. Empirical findings indicate that transformational leadership enhances teachers' openness to pedagogical innovation, willingness to experiment with new instructional methods, and commitment to institutional goals. These effects typically emerge through mechanisms such as self-efficacy, psychological safety, positive organizational climate, and learning-oriented school cultures. Through these mediators, transformational leadership facilitates the institutionalization of innovative practices.

Authentic leadership, in contrast, is rooted in ethical behavior, transparent communication, self-awareness, and the establishment of trust-based relationships with followers. This leadership style strengthens teachers' sense of autonomy, their perceptions of meaningful work, and their psychological well-being. The mechanisms involved include trust, emotional support, ethical climate, and perceptions of justice. Authentic leadership demonstrates a particularly protective function during periods of teacher burnout and professional uncertainty.

Across both leadership styles, emotional intelligence emerges as a significant moderating factor. Leaders with higher emotional awareness strengthen the motivational and innovative outcomes of transformational leadership, while reinforcing the trust-building and ethical sensitivity dimensions of authentic leadership. As such, leadership development programs increasingly integrate emotional and social–emotional competencies.

Comparative Insights and Future Directions

Although both authentic and transformational leadership are associated with positive organizational outcomes, conceptual overlap among measurement instruments complicates efforts to clearly distinguish these constructs. Component-level analyses show that transformational leadership is more closely tied to innovation, institutional engagement, and motivation for change; authentic leadership, however, more strongly influences teacher autonomy, well-being, and perceptions of ethical climate.

Organizational citizenship behaviors and creativity are supported by both leadership styles, yet through distinct pathways:

- In transformational leadership, these outcomes arise through shared vision and institutional identification.
- In authentic leadership, they emerge through trust, honest communication, and ethical norms.

Cultural context is a crucial factor in understanding differential impacts. Transformational leadership may exhibit stronger effects in centralized educational systems, whereas authentic leadership may be more effective in contexts with strong expectations for social support and justice. This aligns with existing findings highlighting culture as a key moderating variable.

Need for Longitudinal and Integrated Approaches

Longitudinal research examining the combined effects of authentic and transformational leadership remains limited. Yet variables such as teacher innovation, school culture, motivation, and well-being evolve over time, making time-sensitive models indispensable. Integrated frameworks evaluating the dynamic interplay between these two leadership styles are therefore essential.

Hybrid models that combine both approaches hold promise for advancing sustainable transformation in education. The visionary and motivational elements of transformational leadership, when merged with the ethical and relational foundations of authentic leadership, can simultaneously enhance innovation and teacher well-being.
Cross-cultural studies would offer valuable insights into how these leadership styles are perceived in diverse governance systems and under varying socio-cultural norms. Research conducted outside Western contexts is especially needed to explain differential effects on teacher creativity, knowledge sharing, and perceived organizational support.

Research Gaps and Mentorship Perspectives

Significant research gaps include:

- Limited studies examining authentic and transformational leadership together,
- Insufficient exploration of mentoring and teacher leadership within leadership development frameworks,
- Measurement tools lacking sufficient discriminatory power to differentiate between the two constructs.

Challenges in mentoring—such as power asymmetries, time constraints, and lack of expertise—highlight the need for deeper investigation. Facilitating factors include collaborative cultures, high motivation, structured professional development, and systematic feedback mechanisms.

Social justice–oriented mentoring models emphasize empowering disadvantaged groups, recognizing diverse identities, and promoting inclusive practices. Emotional intelligence training, conflict management, reflective learning, and multimodal (online–face-to-face) mentoring formats are increasingly emphasized in mentor preparation.

Conclusion

Authentic and transformational leadership represent two influential yet distinct approaches in educational leadership research. Their mechanisms differ-transformational leadership primarily drives innovation and engagement, while authentic leadership strengthens autonomy, trust, and ethical climate-yet they remain complementary in practice.

Recommendations for Future Research

- Expansion of longitudinal, integrative studies addressing both leadership types together,
- Component-level comparative analyses,
- Cross-cultural comparisons in centralized vs. decentralized educational systems,
- Development of culturally sensitive and psychometrically robust measurement tools,
- Increased research on mentoring, teacher leadership, and social justice–centered leadership approaches.

Table 8. Examination of Authentic and Transformational Leadership Styles in Education: Dimensions, Mechanisms, Comparative Analysis, and Future Directions (Based on Relevant Studies Indexed in Scopus)

Theme	Authentic Leadership (AL)	Transformational Leadership (TL)	Integrated/Comparative Perspectives
Core Dimensions	Self-awareness, relational transparency, internalized moral perspective, balanced processing	Idealized influence, inspirational motivation, intellectual stimulation, individualized consideration	Substantial construct overlap; complementary in some contexts
Impact on Teachers	Supports well-being, autonomy, and organizational commitment	Enhances innovation, self-efficacy, job satisfaction, and organizational commitment	Both styles positively influence job satisfaction and organizational citizenship behaviors; TL is generally stronger for innovation
Mediators	Autonomy, socio-emotional competence, organizational trust	Self-efficacy, collective efficacy, psychological safety, organizational culture	Organizational culture mediates both; emotional intelligence and self-efficacy interact more strongly with TL
Measurement	Authentic Leadership Questionnaire (ALQ), contextually adapted	Multifactor Leadership Questionnaire (MLQ-5X), GTLS	High correlation ($\rho \approx .72$); requires detailed, component-level differentiation
Cultural Contexts	Influenced by collectivism, cultural norms, and education policy	Similar, but TL tends to be more robust across contexts	Cross-cultural effects vary; more comparative research is needed
Longitudinal/Integrated Studies	Rare; predominantly cross-sectional	Rare; predominantly cross-sectional	Significant gap in integrated, longitudinal studies

References

Anderson, M. (2017). *Transformational leadership in education*. International Social Science Review, 93(1), 1–13. https://www.jstor.org/stable/10.2307/90012919

Avolio, B. J., & Gardner, W. L. (2005). Authentic leadership development: Getting to the root of positive forms of leadership. *The Leadership Quarterly, 16*(3), 315–338.

Avolio, B. J., & Luthans, F. (2003). *Authentic leadership development*. In K. S. Cameron et al. (Eds.), Positive organizational scholarship (pp. 241–258). Berrett-Koehler.

Ayrancı, E., & Öge, E. (2010). Dönüşümsel liderlik kavramı hakkında önde gelen teoriler ve Türkiye'de kavramı ele alan çalışmalar. *ABMYO Dergisi, 17*, 37–46.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. Freeman.

Bass, B. M. (1985). *Leadership and performance beyond expectations*. Free Press.

Bass, B. M., & Avolio, B. J. (1994). *Improving organizational effectiveness through transformational leadership*. Sage.

Cemaloğlu, N., & Özdemir, M. (2019). *Eğitim yönetimi* (1. bs.). Pegem Akademi.

Çevik, M. S. (2024). A general overview of studies on authentic leadership (1978–2022): A bibliometric analysis. *Bartın University Journal of Faculty of Education, 13*(4), 1083–1110. https://doi.org/10.14686/buefad.1334853

Downton, J. V. (1973). *Rebel leadership: Commitment and charisma in the revolutionary process*. Free Press.

Elrehail, H., et al. (2018). Leadership and innovation in higher education: The mediating role of knowledge sharing. *International Journal of Educational Management, 32*(3), 398–415.

Feng, D. (2016). Authentic leadership and teachers' psychological capital. *Journal of Educational Administration*, 54(5), 1–20.

Gumusluoglu, L., & Ilsev, A. (2009). Transformational leadership and organizational innovation: The roles of internal and external support for innovation. *Journal of Product Innovation Management, 26*(3), 264–277.

House, R. J. (1977). A 1976 theory of charismatic leadership. In J. G. Hunt & L. L. Larson (Eds.), *Leadership: The cutting edge* (pp. 189–207). Southern Illinois University Press.

Hsu, W.-L., Silalahi, A. D. K., Eunike, I. J., & Phuong, D. T. T. (2024). Fostering creativity and performance through authentic leadership in educational institutions. *Cogent Social Sciences, 10*(1), 2425774. https://doi.org/10.1080/23311886.2024.2425774

Kakungulu, S. J. (2024). Transformational leadership in educational institutions. *Eurasian Experiment Journal of Humanities and Social Sciences, 5*(3), 18–22.

Kareem, J., Patrick, H. A., Prabakaran, N., B. V., Tantia, V., M. P. M., & Mukherjee, U. (2023). Transformational educational leaders inspire school educators' commitment. *Frontiers in Education, 8*, 1171513. https://doi.org/10.3389/feduc.2023.1171513

Kernis, M. (2003). Toward a conceptualization of optimal self-esteem. *Psychological Inquiry, 14*(1), 1–26.

Khanyile, S., & De Bruin, G. (2022). Authentic leadership and its impact on organizational outcomes: A systematic review. *Journal of Leadership Studies, 16*(2), 45–67.

Luthans, F., & Youssef, C. M. (2004). Human, social, and now positive psychological capital management. *Organizational Dynamics, 33*(2), 143–160.

Ramos, R. (2025). Transformational leadership in education: Impacts and practices toward enhancing school performance. *Journal of Interdisciplinary Perspectives, 3*(9), 527–536. https://doi.org/10.69569/jip.2025.574

Walumbwa, F. O., Avolio, B. J., Gardner, W. L., Wernsing, T., & Peterson, S. (2008). Authentic leadership: Development and validation of a theory-based measure. *The Leadership Quarterly, 19*(1), 89–126.

Williams, P. (2016). Transformational leadership in education: A review of current practice. *Journal of Education and Learning, 5*(4), 89–95.*

Xie, J., Ahmad, J. B., & Lu, X. (2024a). Authentic leadership: Origins and foundations – A review of literature. *International Journal of Academic Research in Progressive Education and Development, 13*(3). https://doi.org/10.6007/IJARPED/v13-i3/21881

Xie, J., Ahmad, J. B., & Lu, X. (2024b). A comprehensive review of authentic leadership: Previous study, challenge, strategies and practices. *International Journal of Academic Research in Business and Social Sciences, 14*(9). https://doi.org/10.6007/IJARBSS/v14-i9/22802

Yavuz, A. E., & Oğuzhan, Y. S. (2024). Otantik liderlik. In G. Durmuş (Ed.), *Liderlik dinamikleri üzerine araştırmalar I*. Özgür Yayınları. https://doi.org/10.58830/ozgur.pub526.c2236

Developing Higher Education Students' Artificial Intelligence (AI) Literacy for Academic Writing: A Pilot Study in A First-Year Seminar

Aysan Kolahdouzipour
PhD student in Writing and Rhetoric, Department of English, George Mason University, Fairfax, Virginia, USA.
ORCID: https://orcid.org/0009-0008-0775-2148 email: akolahdo@gmu.edu
(Corresponding author)

Derya Kulavuz-Onal
Associate Professor, Department of English, Salisbury University, Salisbury, Maryland, USA.
ORCID: https://orcid.org/0009-0008-8702-6348
email: dxkulavuz-onal@salisbury.edu

Abstract
As generative AI tools like ChatGPT become increasingly integrated into higher education contexts, the need to equip students with critical AI literacy has become urgent. This pilot study investigates how a four-lesson instructional module shaped undergraduate students' understanding of the ethical use, potentials, and limitations of generative AI tools, particularly in relation to academic reading and writing. Implemented in a First-Year Seminar (FYS) course at a U.S. university, the study employed a mixed-methods design, drawing on pre- and post-module AI Literacy Surveys (n=7), a post-semester instructor interview, and the researcher's reflection journal. Findings revealed a notable shift in students' awareness of ethical considerations, collaborative engagement with AI, and development of authorial voice when scaffolded through structured activities. However, these gains did not consistently transfer to independent academic tasks, pointing to the limitations of short-term interventions and the importance of integrating AI literacy instruction within foundational academic writing curricula. As an exploratory pilot study, the findings provide preliminary insight into the pedagogical potentials and constraints of early AI literacy intervention for college-level academic writing, suggesting the need for sustained, curriculum-embedded approaches to cultivate critical and ethical AI engagement among higher education students.
Keywords: AI literacy, higher education, generative AI, ChatGPT, academic writing, student voice, ethics in AI, instructional design, first-year seminar, digital pedagogy

Introduction
In recent years, generative AI (Artificial Intelligence) tools have increasingly transformed not only our professional and personal lives but also our academic lives. With its launch as an open access tool in late 2022, ChatGPT, a generative AI tool, has particularly attracted millions of users around the world, with a significant portion of students and educators exploring its potentials for academic uses (OpenAI, 2023). While such generative AI tools bring new opportunities and affordances, they also raise questions about authorship, authenticity, originality, and ethical engagement with these tools. As Ateriya et al. (2025) pointed out, the use of AI systems to produce content has sparked worries about intellectual-property protection, since these tools may output material that is very similar to existing publications, which in turn heightens the possibility of accidental plagiarism. Such questions are of particular interest to higher education institutions and educators, where students need to learn how to deeply engage with texts, construct evidence-based arguments, conduct research, and create original texts with original arguments.

Due to this increasing use of generative AI tools and the new understandings of their potentials in academic settings, higher education (HE) institutions are now grappling with issues of how to address students' ethical engagement with these tools, as overreliance on these tools create problems of academic integrity, although the tools also have benefits. As Mollick and Mollick (2023) argue, "while AI has the potential to help students learn, its ability to quickly output writing tasks, summarize information, provide outlines, analyze information, and draw conclusions may mean that students will not learn these valuable skills." (p. 3), which raises important concerns about the value of higher education experience. Moreover, despite the increasing use of these tools, many students still lack sufficient knowledge about how they operate and what constitutes ethical uses of AI tools for academic purposes. Educators seeking support with the complexities of AI in education—whether technical, ethical, or conceptual—can refer to institutional policies, international guidelines, or scholarly discussions on the core principles of AI use (Wiese, et al., 2025). Promma, et al., (2025; as cited in Ng., et al. (2022) noted that although AI literacy is vital, research shows that students often lack key competencies in this area, underscoring the pressing need to develop curricula that strengthen these abilities. Therefore, there is now an increasing need in HE settings for educators to "play an active role in teaching students how and when to use AI as they instill best practices in AI-assisted learning" (Mollick & Mollick, 2023, p. 3). However, such studies exploring the development, implementation, and effects of instructional modules

and interventions in HE settings that teach students how to engage with AI tools ethically and responsibly are still scarce in literature.

Designing a curriculum that fully addresses the multifaceted skills and knowledge outlined in AI education frameworks is a highly challenging task as there is no universal model that fits all contexts (Wiese et al., 2025). Even when focusing narrowly on AI ethics literacy, instructors must navigate complex factors such as students' lived experiences and external institutional or societal pressures (Dai et al., 2023). Moreover, the issue is not only about AI literacy but also about academic literacy. As reading and writing are central to most university curricula, it is important to design learning opportunities to develop students' ability to critically engage with AI in these areas. This means going beyond simplistic 'do not use ChatGPT' policies to instead foster meaningful conversations and activities with students about what constitutes appropriate use, what limitations AI-generated content has, and how students can responsibly integrate these tools into their learning. Students need to learn the technical aspects of these tools, how to evaluate these tools, and what constitutes academic integrity, all at the same time.

Departing from these gaps and needs in the literature, we designed, implemented, and piloted a 4-lesson AI-literacy-focused instructional module, focusing particularly on AI use in academic writing, in a first-year seminar of a HE institution in the United States. This pilot study explored how such an intentional pedagogical intervention can shape first-year students' AI literacy in relation to academic writing tasks. More specifically, this pilot study aimed at addressing the following research question: "How does participation in an AI-literacy-focused instructional module shape undergraduate students' understanding of the ethical uses, potentials, and limitations of generative AI tools like ChatGPT in relation to academic writing tasks?"

Literature Review

Generative AI in Higher Education

The emergence of generative AI tools has sparked intense interest in higher education settings. Particularly ChatGPT, the tool that was used in this pilot study, quickly became popular among higher education (HE) students across disciplines. Recent studies have shown that HE students use it regularly for various academic tasks, such as coding or lesson planning depending on their major, but also for a variety of writing tasks from brainstorming to editing regardless of students' specific majors (Onal, Kulavuz-Onal, & Childers, 2025). The widespread adoption of ChatGPT by HE students is because of not only its tool-specific versatility, efficiency, and ease-of-use as perceived by students (Holland & Ciachir, 2024; Fauzi, et al. 2023), but also social norms and campus infrastructure (Agyare et al., 2025) and individual differences among students, such as students' positive attitudes towards AI and digital literacy and confidence (Kim, Ham, and Lee, 2024). Moreover, students perceive ChatGPT to be useful and sometimes treat it as a 'private tutor' for understanding complex academic tasks, such as reading lengthy academic texts (Das & Madhusudan, 2024), translating and paraphrasing (Gupta, et al., 2022; Kim et al., 2025), solving problems, generating ideas, and saving time while managing multiple tasks at a time (Onal, et al., 2025).

Despite its perceived usefulness, many students also report concerns with the use of AI tools for academic tasks. These concerns particularly center on academic integrity violations and 'hallucinated' facts created by AI tools that lead to spreading misinformation and causing mistrust in AI-generated material (Črček & Patekar, 2023; Kim, et al., 2025). There have also been raising concerns about the AI tools' environmental impact among students, although their concerns of immediate educational outcomes still surpassed such environmental concerns (Zhao, et al., 2024).

Academic Writing in the Age of AI

Artificial Intelligence (AI) tools, particularly ChatGPT, have become a supportive learning tool for HE students especially in academic writing. Recent studies show that students are using ChatGPT for a variety of academic writing tasks, benefiting from its capabilities to enhance grammatical accuracy and writing quality, improve coherence, streamline idea development and provide translation and paraphrasing support (Amrullah et al., 2024; Onal et al., 2025; Wang, 2024). Moreover, the tool supports learners' efficiency in writing by providing immediate feedback, enabling better vocabulary use, and facilitating structured outlines and topic exploration (Amrullah et al., 2024; Wang, 2024).

Studies demonstrated that students generally perceive AI tools as versatile partners in the writing process. In Kim, Yu, Detrick and Li (2025) study, students described AI as a "multi-tasking writing assistant," a "virtual tutor," and a "digital peer," capable of supporting ideation, drafting, revising, and proofreading tasks. Similarly, Črček and Patekar (2023) found that students most commonly used ChatGPT to generate ideas, summarize texts, and check grammar and spelling. Students appreciated AI's capacity to scaffold writing by offering immediate feedback and suggestions, positioning it as a time-saving and motivational resource. Such positive experiences with AI tools in academic writing tasks also showed positive correlation with students' level of writing self-efficacy (Bouzar. et al, 2024).

Studies also reported students' use of AI tools such as ChatGPT for academic reading purposes in connection with their writing tasks. Onal et al., (2025) found that it has been increasingly common for students to use ChatGPT to summarize long or complex readings assigned as homework or check their understanding of texts, especially in source-based writing tasks. Other studies also suggest that students get assistance from ChatGPT for rephrasing text and summarizing readings (Zhao, et al. 2024).

In addition to benefits and increasing adoption of AI tools such as ChatGPT for academic reading and writing tasks, students also report hesitancy and caution in their use. Cummings, Monroe, and Watkins (2024), examined that many first-year students still preferred to develop their own ideas and voice, expressing reservations about the reliability of AI outputs. According to their study, when given the choice, students often limited their use of AI to early drafting stages, preferring to revise and synthesize ideas independently as their writing progressed. Similarly, Tarchi, Fidalgo, and Brante (2024) observed that HE students in their study with source-based writing tasks using ChatGPT, rarely copied AI-generated content verbatim, instead using it selectively. While some viewed ChatGPT as helpful for paraphrasing or identifying relevant themes, others doubted its credibility.

Artificial Intelligence (AI) Literacy

With the rapid integration of AI into all aspects of our lives, there is now an urgent need for educating the public to engage with AI tools in an informed and critical way. The concept of *AI literacy* has emerged in response to this need, aiming to equip learners with the knowledge and skills necessary to understand, evaluate, and use AI systems ethically and effectively. One AI Literacy framework developed by Mills et al. (2024) defines AI literacy as "the knowledge and skills that enable people to critically understand, evaluate, and use AI systems and tools to safely and effectively participate in an increasingly digital world" (p. 4). They assert that "understanding and evaluating AI is critical to making informed decisions about if and how to use AI in different contexts" (p. 5). The framework is built around three interrelated Modes of Engagement**:** Understanding**,** Evaluating**,** and Using AI. These components are not linear stages but concurrent, mutually reinforcing ways of engaging with AI systems. Each of these modes is operationalized through six AI Literacy Practices (algorithmic thinking, data analysis and inference, data privacy and security, digital communication and expression, ethics and impact, and information and mis/disinformation) and emphasizes two core values: *centering human judgment* and *centering justice*. For example, in the practice of Algorithmic Thinking, students are expected to break down complex problems and understand how algorithms automate decisions. In Ethics and Impact, learners question how datasets and algorithmic design can reproduce bias and cause harm. It is intended that learners' understanding of the workings of these systems and the likely consequences of unethical uses that may result in dissemination of misinformation should be understood by learners and evaluated carefully for them to use them effectively and ethically. As such, to develop AI literacy, students need to engage in such activities and practices, preferably with the guidance of educators in their unique learning contexts.

According to Almatrafi, Johri, and Lee (2024) and Ng et. al. (2021), AI literacy frameworks generally integrate six common constructs: *Recognize*, referring to understanding and noticing whether or not AI was used in the creation of an output; *Know and Understand*, referring to acquiring foundational knowledge of how AI systems function, including concepts like data, algorithms, and machine learning processes; *Use and Apply*, which involves interacting with AI systems, using them purposefully, and applying them in diverse contexts for learning, communication, or problem-solving; *Evaluate*, which refers to critically assessing the outputs of AI systems, understanding their limitations and potential biases, and determining their appropriateness for specific tasks; *Create*, referring to designing, modifying, or developing AI systems or tools, including the ability to train models or customize outputs through prompting or coding; and *Navigate Ethically*, which encompasses awareness of and engagement with the ethical, social, and cultural implications of AI, such as fairness, transparency, privacy, and accountability. Almatrafi et al. (2024) also found that many studies emphasize the necessity of learners engaging in critical evaluation of AI tools and using and applying AI as fundamental components of AI literacy frameworks.

Instructional Interventions for Teaching AI and AI Literacy Skills

Recent empirical and theoretical work has examined instructional interventions to foster AI literacy in both K–12 and higher education. Systematic reviews show that K–12 AI literacy initiatives generally follow two paths: direct learning experiences aimed at technical, conceptual, and applied skills, and curriculum frameworks that provide overarching design principles (Casal-Otero et al., 2023). The "AI4K12" initiative's "five big ideas" has informed many such curricula, offering conceptual anchors for embedding AI into STEM and interdisciplinary programs (Gu & Ericson, 2025).

Though still in its infancy, research on exploring the effects of structured interventions to help students develop AI literacy that focus not only on technical but also evaluative and ethical competencies has also been growing. The results of these studies suggest engaging learners in hands-on contextualized learning tasks where they directly engage and interact with AI tools. In their systematic review, Casal-Otero et al. (2023) identified that in K-12 settings, efforts of AI literacy development generally included activities where students learned to recognize AI-generated artifacts, understanding how AI works, and critically evaluating how AI impacts our lives.

Many interventions adopt task-based or design-based learning models, providing opportunities for students to create AI-driven projects. For example, Akman (2025) investigated the effects of AI-assisted video and graphic design projects on college undergraduate students' AI literacy. Both groups showed significant improvements in AI literacy and attitudes toward AI, with the video design group demonstrating a greater increase than the graphic design group. Similar evidence comes from middle school programs such as the DAILy workshop (Lee et al., 2021) and the ActiveAI goal-based scenario app (Tseng, et al., 2024), which integrate project-based learning with ethical considerations. The AI Audit card game (Ali, Kumar, & Breazeal, 2023) further illustrates how gamified simulations can prompt learners to design, critique, and reconsider AI systems, fostering awareness of potential societal harms. These findings suggest that authentic, creative, task-based and production-oriented activities may enhance not only technical understanding but also evaluative and ethical engagement.

Specifically, in teaching AI literacy in relation to academic writing in HE, studies show similar recommendations for critical engagement with AI-generated outputs. For example, Van Niekerk, Delport, and Sutherland (2025) conducted an active learning intervention in which students generated academic essays using ChatGPT and then critically evaluated them based on pre-determined academic writing criteria. Initially impressed by the tool's ease of use and fluent output, students later became more skeptical after recognizing limitations such as fabricated references, lack of critical insight, and superficial argumentation. The intervention successfully reduced students' behavioral intention to rely on ChatGPT uncritically and fostered metacognitive engagement by prompting students to reflect on the tool's appropriate and inappropriate uses in academic writing. Similarly, Agyare, et al. (2025) conducted a cross-national study to investigate college physics students' perceptions and usage of ChatGPT across four universities in Ghana, Jordan, and the United States. They found that ethical concerns negatively moderated students' behavioral intentions in using this technology, indicating that students who were more ethically cautious were less likely to use ChatGPT despite favorable intentions, which supports integrating ethical AI literacy into instructional design. Moreover, Yang, Huang, and Liu (2025) found that when instructors encourage students to reflect critically on the advantages and risks of AI tools, learners begin to treat AI as a collaborative resource rather than an authoritative one. Their study highlights that effective integration of generative AI requires not only tool adoption but also shifts in pedagogical mindsets, where students and teachers co-construct norms for responsible AI use.

All in all, studies on instructional interventions to teach AI literacy collectively recommend embedding contextualized, participatory and active engagement with AI tools, explicit ethics discussions, discipline-specific guidelines, and reflective activities in the classroom to move students from passive consumption toward critical, strategic AI use.

Voice in Academic Writing

In academic writing, voice is a complex and debated concept, yet Elbow shows it remains central to how students and teachers understand written communication. He notes that readers often experience writing as if it carries an "audible voice" – a quality that makes language clearer and more engaging because "when readers hear a voice in a piece of writing, they are often more drawn to read it – and that audible voice often makes the words easier to understand" (Elbow, 2007, p. 7). Elbow also clarifies that the term voice contains multiple dimensions, explaining that he has previously distinguished five senses of voice: audible voice, dramatic voice, recognizable or distinctive voice, voice with authority, and resonant voice (Elbow, 2007, p. 12). Each of these describes a different way writers project presence, stance, or personality through language. Elbow further emphasizes why attention to voice matters for

developing academic writers: students instinctively talk about voice, teachers continue to reference it, and writing that carries a sense of voice is typically easier to follow and more rhetorically effective because "when they do [write with voice], their words are more effective at carrying meaning" (Elbow, 2007, p. 9). His discussion also shows that attention to voice helps writers shape how they are perceived, since qualities such as sincerity, authority, or resonance influence how readers interpret ideas and respond to arguments. For first-year college students – who are still forming their academic identities and learning to navigate unfamiliar discourse expectations – Elbow's work demonstrates that understanding voice and learning to control it can support clearer communication, stronger ethos, and greater confidence as they enter academic conversations.

Several studies have explored how to help higher education students develop their voice in academic writing. Vengadasalam (2020) emphasizes that voice is "how each student contributes to scholarship" and that developing it involves "struggle, risk-taking, and reconstruction," (p. 13) making voice an important marker of academic success. Because students' voices can be silenced by academic norms or feelings of marginalization, she argues that instructors must intentionally create classroom spaces where students feel safe, recognized, and able to assert their perspectives. To guide this work, she proposes the S.E.A. framework – scaffolding, empowerment, and awareness. Empowerment comes from creative, interdisciplinary writing prompts that encourage students to take innovative positions. Awareness grows through critical questioning of texts and disciplinary conventions. Scaffolding supports voice development by structuring the writing process and modeling voiced writing through outlining, drafting, revising, and peer review. Together, these practices help students develop confident and original academic voices. Similarly, Healey, Matthews, and Cook-Sather (2020) emphasize that developing a strong voice in academic writing involves intentionally presenting one's identity, values, lived experience, and standpoint. They explain that an author's voice conveys identity, experiences, values, and perspectives, allowing writers to assert authority while resisting impersonal, jargon-laden prose. Using examples such as Alise de Bie's personal narrative and Tara Yosso's authoritative assertions, the authors illustrate how voice helps writers situate themselves within scholarly conversations and make their ideas more hearable.

Significance of the Present Study

Across AI literacy scholarship, several recurring constructs emerge despite differences in terminology. Frameworks such as Mills et al. (2024), Ng et al. (2021), and Almatrafi et al. (2024) consistently emphasize three interrelated dimensions of AI literacy: (1) understanding how AI systems function, (2) critically evaluating AI-generated outputs and their ethical implications, and (3) using AI tools strategically and responsibly in domain-specific contexts. These dimensions suggest that effective AI literacy instruction should not focus solely on technical knowledge but should also engage learners in critical evaluation and purposeful application of AI tools within authentic tasks.

In the context of academic writing, these dimensions intersect with the development of personal voice, where students must learn to use AI tools as supportive collaborators rather than replacements for their own ideas and rhetorical choices. Drawing on these converging perspectives, the present study conceptualizes AI literacy in academic writing as the ability to critically evaluate AI outputs, ethically apply AI tools, and integrate AI assistance while maintaining personal authorial voice

Developing voice in academic writing is not isolated from understanding the constraints of genres of writing, by which students gain insight into how to align their writing with academic standards while negotiating their individual voice. In this effort, AI tools offer additional support for students to navigate academic conventions. However, as AI literacy frameworks also emphasize, as discussed above, students need opportunities of active engagement to understand the role of personal voice in academic writing, and how AI tools can reinforce, not impede, their authorial identity. To our knowledge, there have been no instructional intervention in HE and Generative AI literature that explored the potential of the design of such an instructional intervention to help students develop and integrate their voice into academic writing through developing their AI literacy and learning to use AI tools responsibly and ethically. This pilot study aims to provide perspectives into how this could be achieved in an HE classroom with first-year undergraduate students.

Method

Setting and Context

This study was conducted in a First-Year Seminar (FYS) course at a regional university in the US that focuses on undergraduate education. FYS is a 100-level general education requirement typically taken in the first year alongside First-Year Composition (FYC). Students usually enroll in FYC during the fall semester and FYS in the spring, though

the sequence may vary. Both courses support students' academic literacy and critical thinking, often working in tandem to develop foundational skills in writing, inquiry, and reflection. In addition to its focus on developing students' academic literacy, this FYS course was also chosen as the setting for this pilot study because it explores language-related issues across different disciplines, such as language matters in AI and technology, among others.

Instructional Module Overview

In order to purposefully align with the content of this First-Year Seminar course, this instructional module was implemented during the *Language Matters in Technology* unit. Spanning two weeks, the module consisted of four lessons, each lasting 75 minutes and held twice a week. Designed to align with the course's emphasis on language and social relevance, the module explored the language of AI and the intersection of personal voice and artificial intelligence in academic writing.

Conceptual Framework for the Instructional Module

The design of the instructional module was informed by converging AI literacy frameworks that emphasize three core dimensions of engagement with AI systems: understanding AI functionality, critically evaluating AI-generated outputs, and using AI tools responsibly and ethically in authentic contexts (Almatrafi et al., 2024); Mills et al., 2024; Ng et al., 2021. These principles were integrated with scholarship on personal voice in academic writing, which emphasizes the importance of helping students maintain ownership, stance, and rhetorical identity when engaging with external sources or tools (Elbow, 2007; Hyland, 2002; Ivanič, 1998). Together, these perspectives guided both the instructional design and the measurement of learning outcomes in this study.

With the increasing uses of AI tools in Higher Education institutions, professors in FYS and FYC courses in this university also noted significant increases in students' AI use without understanding how AI works and how to use them responsibly and ethically. Since these courses are taken by students in their first years, and have a focus on understanding the demands of academic life, especially as regards to academic writing and research, it was determined that teaching and guiding students about AI tools in academic writing and research may be a useful addition to these courses. Therefore, in developing and piloting this instructional module, the goal was to improve students' understanding of ethical uses of AI tools in their academic writing in ways that raise their awareness of personal voice in writing and how to adopt AI tools to support one's writing rather than replacing it. With these goals in mind, the four lessons in this module were organized in a consecutive manner as shown in the flowchart in Figure 1 below:

Figure 1

Flowchart of the Two-Week AI Literacy Instructional Module

Lesson 1: Student Voice in Writing
This lesson focused on the importance of personal voice in academic writing. Students explored why voice matters and examined strategies for expressing their personal voice in their writing.

Lesson 2: Introduction to Large Language Models and AI Functionality
This lesson introduced students to large language models and how they function. Emphasis was placed on maintaining a personal voice while engaging with AI-generated content. As part of this lesson, students interacted with AI outputs and reflected on ways to engage with and personalize them.

Lesson 3: Prompts and Practical Applications
In this lesson, first the key concepts were reviewed, including student voice and the distinctions between human and AI-generated writing. Students were introduced the idea of creating effective prompts to support personal and academic goals in an ethical way. As part of this lesson, students also practiced creating prompts for different scenarios, focusing on how to guide AI in producing academically ethical content.

Lesson 4: Practice and Wrap-Up
In this lesson, students practiced identifying ethical versus unethical uses of AI tools through various scenarios and refined their prompt development. The lesson concluded with a review and reflective discussion on the main takeaways.

Collectively, following the existing AI Literacy frameworks' components of recognizing, knowing and understanding, using and applying, creating, evaluating, and navigating ethically, the four lessons were designed to move students from understanding AI systems, to critically evaluating AI outputs, and finally to strategically using AI tools in ways that preserve their own authorial voice in academic writing, reflecting the progression suggested by AI literacy frameworks.

Participants
Participants were first-year undergraduate students all enrolled in the same First Year Seminar (FYS) course. A total of 18 students were registered in the course roster, including one student never attending any of the modules. The class population included 9 female and 9 male participants all of whom were 18 years of age or older. This population was chosen because first-year students are often new to higher education contexts and may face challenges navigating academic reading and writing demands. As mentioned previously, this course was selected due to its thematic alignment with language and technology, including content related to AI tools and digital literacy, making it an appropriate context for the intervention.

Data Collection and Analysis
Data for this study came from three sources: a Perceived AI Literacy Survey that was collected before and after the instructional module, the first author's (also the module instructor) researcher reflection journal, and an end-of-semester interview with the FYS course instructor. All participants completed informed consent forms prior to data collection.

Perceived AI Literacy Survey
An AI Literacy Survey was created by reviewing the previous research and through expert reviews prior to the study to explore students' perceptions and beliefs about their AI literacy, rather than their actual AI Literacy performance. The survey then was distributed online to students through Qualtrics before the instructional module and right after the final lesson in the module.

The finalized survey consisted of 26 items asking students their perceptions on a 5-point Likert-scale, 1 being 'strongly disagree', and 5 being 'strongly agree'. Internal consistency for the overall AI literacy survey was assessed using Cronbach's alpha. The 26-item scale showed acceptable reliability, $\alpha = .88$, indicating that the items collectively measured a cohesive construct. The items in the survey were also grouped into 4 different categories reflecting key dimensions of AI literacy identified in prior frameworks, including use and application of AI tools, ethical awareness, critical collaboration with AI in writing, and perceptions of AI's role in academic success (Almatrafi et al., 2024; Mills et al., 2024; Ng et al., 2021). These categories were used to support analysis and interpretation of the results, as shown in Table 1.

Table 1. *Perceived AI Literacy Survey Components*

Category	Number of Items	Sample item	MPS*
Use and Application of AI Tools in Academic Contexts	7	*I believe AI tools can help me improve my writing skills.*	35
Awareness of Ethical Considerations	4	*I am aware of the unethical uses (i.e., misuses) of AI tools for academic purposes.*	20
Collaborative Writing with AI Tools and Voice	7	*I know how to add my personal voice to AI-generated content in a way that maintains my academic integrity.*	35
AI's Impact on Student Success and Access	8	*I believe AI tools can help me improve my writing skills.*	40

*MPS: Maximum Possible Score

In this survey, *Use and Application of AI Tools in Academic Contexts* (7 items; MPS = 35) assessed students' perceived familiarity with generative AI and confidence using tools like ChatGPT to support academic reading and writing (e.g., drafting, idea generation, comprehension support), including perceived ability to write effective prompts and review/revise AI outputs for errors. *Awareneess of Ethical Considerations* (4 items; MPS = 20) category aimed to capture students' perceived awareness of ethical concerns and misuses of AI in academic work (e.g., plagiarism, bias), as well as their perceived understanding that responsible use involves actively interacting with AI and revising outputs rather than relying on a first response. *Collaborative Writing with AI tools and Voice* (7 items; MPS = 35) focused on understanding students' perceived ability to work with AI as a writing partner while maintaining their writing voice, style, authorship and ownership of their work (e.g., differentiating AI vs. human text, using AI output as a starting point, revising and reshaping AI-generated content, and integrating their personal academic voice in ways that align with integrity). Finally, *AI's Impact on Student Success and Access* category explored students' perceived beliefs about AI's usefulness for academic success and access, including whether AI can improve writing quality, increase efficiency, support comprehension, expand access to resources, and help address language- or writing-related challenges.

In this study, only data from the students who provided informed consent and who completed both the pre- and post-module surveys were included. A total of 7 students completed both surveys, although all 18 students participated at least three of the four lessons in this instructional module. Scores from individual items were summed within their respective categories to produce an overall score for each student in each of the survey's five categories. Because of the small-scale nature of this study, the data with students' pre- and post-module survey scores were analyzed through descriptive statistics.

Interview with the Course Instructor

In addition to student data, a one-hour semi-structured interview was conducted with the course instructor toward the end of the semester. This interview explored the instructor's perspective on students' engagement with AI tools and any observable changes in their writing practices or written products over the course of the semester, through interview questions such as "Have you noticed any shifts in how students are using AI tools since the AI literacy modules were introduced? In what way?", and "What else do you think could be done to raise students' awareness of effective and ethical use of AI tools for academic purposes?" among others.

Qualitative data obtained through this semi-structured interview with the course instructor were then analyzed following the qualitative thematic analysis techniques through coding, theming, and categorizing the transcribed interview data (Braun & Clarke, 2006; Saldana, 2021). After the interview was transcribed, the transcription was coded by both researchers. These initial codes were then categorized into larger themes in response to the research question.

Researcher Reflection Journal

This reflection journal was written retrospectively by the first author after the completion of the instruction on the four-lesson AI literacy module in the FYS course. While informal notes and impressions were taken during and immediately after each lesson, the full journal was composed at the end of the module. This allowed for a broader

view of the overall progression – particularly how students' understanding of voice, AI functionality, and ethical writing practices evolved over time.

The reflections focused on the first author's observations and experiences while teaching each lesson, with particular attention on classroom activities including in-class writing samples, peer-sharing discussions, and instructional materials, as well as her evaluations on which activities most effectively supported student learning and engagement, and how student responses challenged or confirmed the pedagogical intentions behind each lesson. Special attention was given to moments when students demonstrated increased confidence, questioned AI-generated content more critically, or engaged in nuanced ethical reasoning.

Writing retrospectively provided space to synthesize immediate observations with long-term outcomes. It also helped surface areas for improvement, such as the need to adjust scenarios to better match students' academic backgrounds and writing skills. Overall, the journal served as both a pedagogical self-assessment and a qualitative record of how students negotiated voice, authorship, and responsibility in AI-supported writing environments to triangulate the quantitative data obtained in the pre- and post-module surveys.

Ethical Considerations and Researcher Positionality

This study was approved by the Institutional Review Board of the institution where it took place. Additionally, several steps were taken to ensure ethical integrity and data reliability. First of all, all the data collection procedures (including the surveys, and the teaching of the instructional module) was completed by the first author who maintained a neutral role in the classroom. As this was the first time meeting the students, the first author was positioned solely as a guest instructor facilitating the instructional module, with no involvement in grading or prior relationships with the participants. This separation helped reduce bias and increase the trustworthiness of the data collected. All participation was voluntary, and students were informed that they could withdraw at any time without academic consequences. On the other hand, although the researcher was not the course instructor for this particular course, she was an instructor for First-Year Composition courses. Therefore, she was knowledgeable and experienced about the pedagogies of teaching and the needs of first-year college students.

Moreover, the actual course instructor was not present at any time during the course of the two weeks when the instructional module was implemented. This allowed for neutral participation by the students during the lessons, free of fear of their grades being affected by their participation or lack thereof.

The surveys did not ask for any identifiable information, such as students' names or majors. Each participant was assigned a unique identification number to ensure confidentiality and to match pre- and post-surveys. Only the first author had access to the identification key. Survey responses remained confidential and were accessible only to the student researcher and the principal investigator. All data were de-identified prior to analysis. The second author only had access to this de-identified data.

All students in the course engaged in the instructional module as part of regular class activities; however, participation in the survey was entirely voluntary. Both pre- and post- surveys were collected by the first author (aka "the researcher"), who designed and facilitated the instructional module but had no role in grading or academic evaluation of the students. The course instructor was not present during survey collection and remained unaware of which students participated.

Results

In this study, a 4-lesson instructional module was designed to teach students how to use Generative AI tools, specifically ChatGPT, in academic writing in ethical and responsible ways considering their own voice. In order to understand the usefulness of this module, it was implemented with first-year college students in a First Year Seminar (FYS), and collected data to answer the following research question: "How does participation in an AI-literacy-focused instructional module shape undergraduate students' understanding of the ethical uses, potentials, and limitations of generative AI tools like ChatGPT in relation to writing tasks?" To answer this question, data were collected from three sources: pre- and post-module Perceived AI Literacy Surveys, a post-semester interview with the course instructor was analyzed, and the researcher's reflection journal as the module instructor. Because the survey assessed perceived AI literacy rather than performance-based competencies, the findings should be interpreted as students' reported confidence and awareness rather than direct evidence of skill development.

All 18 students in this class attended at least three of the four lessons in the instructional module. However, 7 of them completed the AI-literacy survey both before and after the module. Therefore, Table 2 below presents only these students' scores and the descriptive statistics on the Perceived AI literacy Survey before and after the instructional module, based on their self-reported data.

Table 2. *Pre- and Post-Module Perceived AI Literacy Survey Scores across Four Categories*

St ID #	Use and Application of AI Tools in Academic Contexts (MPS*: 35)		Awareness of Ethical Considerations (MPS*: 20)		Collaborative Writing with AI Tools and Voice (MPS*: 35)		AI's Impact on Student Success and Access (MPS*: 40)	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post
1	26	28	16	16	23	28	32	27
4	26	35	20	20	20	35	25	40
5	27	35	17	20	26	34	34	40
6	26	35	18	20	30	35	34	40
13	23	25	13	15	21	27	25	28
17	27	35	15	20	28	33	32	40
18	21	26	13	16	15	26	26	30
M	25.14	31.29	16	18.14	23.29	31.14	29.71	35
SD	2.10	4.37	2.39	2.17	4.77	3.68	3.88	5.83

*MPS: Maximum Possible Score

Given the small number of students who completed both surveys (n = 7), inferential statistical analyses were not computed; therefore, descriptive statistics were used to explore patterns in students' self-reported perceptions following the module. Overall, Table 2 above suggests positive shifts in students' self-reported perceptions following the module, with increases in mean scores in all four categories, although they should be interpreted as exploratory and descriptive rather than conclusive. The largest increase appears in *Collaborative Writing with AI Tools and Voice* category (Pre $M = 23.29$, $SD = 4.77$; Post $M = 31.14$, $SD = 3.68$), indicating that students left the module feeling more confident about revising AI-generated text, maintaining ownership, and integrating their voice. Also, the decrease in SD suggests students became more consistent in these perceptions after the instructional module. *Use and Application of AI Tools in Academic Contexts* also shows substantial perceived shifts (Pre $M = 25.14$, $SD = 2.10$; Post $M = 31.29$, $SD = 4.37$), indicating an increase in perceived competence using AI for academic tasks and prompting. On the other hand, the increase in SD in this category implies more variation in how strongly students reported growth and that they perceived varying degrees of benefits of the instructional module. *AI's Impact on Student Success and Access* showed perceived increases as well (Pre $M = 29.71$, $SD = 3.88$; Post $M = 35.00$, $SD = 5.83$), suggesting stronger beliefs in AI's potential to support achievement and access, as opposed to pre-module beliefs, though the larger post-module SD signals that students diverged more in how beneficial they believed AI to be (partly reflected in one student reporting a decline (e.g., Student (St) #1). Finally, *Awareness of Ethical Considerations* shows a smaller but still positive mean increase (Pre $M = 16.00$, $SD = 2.39$; Post $M = 18.14$, $SD = 2.17$), suggesting modest growth in students' perceived ethical awareness with slightly reduced variability; however, some students showed no change at the top end (e.g., students already at the MPS). Although these observed increases are not statistically significant because of the small sample that represents only one class at one university, they may indicate possible shifts in students' perceptions of their AI literacy, suggesting that the module most strongly may have contributed to students' perceived ability to collaborate with Generative AI while maintaining voice and control, while also improving perceived use/application skills and (for 6 out of 7 students) perceived impact on success/access, with more modest but still meaningful gains in perceived ethical awareness.

The quantitative perceived AI literacy data collected before and after the module were also triangulated with the researcher's (the first author's) reflective journal and the end-of-course interview with the course instructor to gain qualitative insights into how the module might have shaped students' understanding and experiences of AI in education and academic writing during the module and throughout the semester. In the sections below, a cross-analysis of quantitative and qualitative findings is reported to understand patterns that emerged in the data that can be attributed to students' perceived shifts (or lack thereof) at the end of the module, pertaining to the focus areas in the research

question: understanding of the ethical uses, potential, and limitations of generative AI tools such as ChatGPT in relation to academic writing tasks.

Understanding the Ethical Uses of AI Tools

In the Perceived AI Literacy Survey the *Awareness of Ethical Considerations* category primarily targeted students' overall beliefs and confidence in understanding the overall ethical issues with and misuses of AI tools in academic work. This category showed small to moderate post-module shit's for 5 out of 7 students, with St #1 reporting the same score and St #4 reporting the MPS of 20 in both pre- and post-module surveys. 4 out of 7 students reported the MPS of 20 in the post-module survey, with St #17 reporting the highest increase (5 points) in their perceived awareness of the ethical considerations at the end of the module. These moderate shift's suggest that students perceived an increased awareness of responsible and ethical AI use after the module, though they were not completely unaware of them pre-module. Similar observations were also noted in the researcher's reflection journal as she noticed thoughtful ethical reasoning emerging during class activities, particularly in Lesson 4 when students evaluated ethical/unethical practices in imaginary scenarios. According to her, "Students were highly engaged during the scenario-based discussions, often debating what should be considered ethical" and she found that students' responses were "thoughtful and nuanced [...] especially when they challenged the examples I considered ethical."

The course instructor during the end-of-semester interview, however, expressed no meaningful change in students' understanding of ethical AI use in general. In reference to a task where students were supposed to conduct interviews with selected participants, the instructor observed: "Most of them appeared to just create the quotes and everything using AI […] it was obvious that, that was not something that, you know, a nurse would say [in an interview]." This observation complicates the positive shifts reported in the post-module survey results by the students. Apparently, for that task, students used AI-generated interview data rather than conducting authentic interviews themselves. He explained that he had "not noticed a shift per se […] those who were using AI continue to use it […] but they're just using AI to generate everything," indicating that, in his view, students' use of AI in graded work remained largely habitual and insufficiently critical later in the semester. Rather than critically revising or interacting with AI-generated content – as seen during hands-on classroom tasks in the instructional module – the instructor perceived a pattern of overreliance on AI and minimal critical intervention. Building on this concern, the instructor argued that the definition of ethical AI use may be relative and task-based and should be determined by placing learning at the center of the task at hand; as he framed it, students should ask themselves questions such as "is it helping [me] learn?" He emphasized that anytime students learn something from interacting with AI, it should be considered a gain, but he also suggested that clearer ongoing guidance is needed beyond one-time intervention, including "a program-wide course on this [AI use] to set the ground rules".

This contrast between the observed growing awareness of ethical AI use during the instructional module (as reflected in the researcher's journal and post-module survey scores) and the continuance of ethically questionable AI practices later in the semester suggests that students' ethical AI awareness was stronger when course tasks explicitly scaffolded and focused on reflective, task-based, learning-centered AI use (e.g., with prompt chaining and revision) under the guidance of an instructor. However, these skills did not adequately transfer to new assignments and tasks beyond the module. Moreover, these findings also suggest that ethical decision-making about the use of Generative AI tools in academic tasks might be task-dependent and depend on students' attitudes towards and understandings of academic integrity. As such, students might need further guidance and focused instructional scaffolding across different assignments to develop more ethical and responsible practices with AI tools.

Understanding AI's Potentials and Limitations in Academic Writing

The main goals of the instructional module in this pilot study were to raise students' awareness of AI's potentials and limitations when using them in academic contexts, particularly in academic writing. As such, several activities throughout the module, particularly in Lesson 2, engaged students in critically analyzing AI-generated texts and discussing how such texts may (or may not) meet expectations for academic writing. The researcher's reflective journal noted that, during the module, students "correctly identified the AI-generated text by noticing its lack of depth, vague tone, and absence of real-world examples." Students were able to distinguish AI-generated writing from human-written academic writing, and articulate specific textual features that signaled limitations in AI authorship and output in meeting academic writing expectations.

Consistent with these classroom observations, the AI Literacy Survey results showed overall growth in students' perceived ability to use AI tools purposefully in academic contexts. As shown in Table 2, students' scores increased

in *Use and Application of AI Tools in Academic Contexts* (Pre: *M* = 25.14, *SD* = 2.10; Post: *M* = 31.29, *SD* = 4.37), with all seven students reporting post-module increases (e.g., St #4: 26 – 35; St #6: 26 – 35; St #18: 21 – 26). Together, the survey data and the researcher's observations indicate that students moved beyond viewing AI output as inherently reliable and began approaching it with greater attention to its quality, specificity, and usefulness for academic writing tasks.

In addition to recognizing limitations in AI-generated texts, during the module, students practiced revising AI-generated paragraphs to reflect their own ideas and writing voice. For example, Lesson 1 focused on teaching students the role of voice in academic writing and students practiced rewriting statements by using voice strategies such as adjusting tone, stance and sentence structure. The researcher acknowledged in her journal that "this lesson was one of the most productive and engaging" as it helped students understand that "their voice is what personalizes their writing". During the activity, she observed that "As students shared their rewritten versions aloud, I could see from their expressions that they were beginning to understand and appreciate the value of their own voice." Students also expressed in their exit ticket quotes a growing sense of confidence in asserting their opinions and recognized the importance of voice in making their writing more authentic and impactful. Several noted a desire to move beyond bland or impersonal writing by incorporating more emotion, clarity, and individuality in their work. As she noted, "Watching them do this so successfully – and seeing how each version became distinct and personal – was one of the highlights of the lesson." Furthermore, in Lesson 2, despite beginning with the same prompt and output, students applied voice strategies introduced in Lesson 1 – such as modifying tone, integrating specific examples, and asserting perspective – to revise AI-generated outputs. The researcher observed that this activity resulted in distinctly different texts at the end of the lesson and it contributed to students' growing ability to adapt generic AI outputs into personalized academic writing, indicating not only technical competence but also a growing awareness of authorship and individuality in academic writing. This also affirms the module's goal of supporting student agency in AI-supported environments. According to the researcher, students articulated intentions to "be more confident in expressing their opinions" and "make their voice louder and clearer." These student comments reflect a shift from passive AI consumption to active, author-driven engagement, which in turn shows their growing ability to turn the limitations of Generative AI tools into potential collaborations with them.

These observations are also aligned with the survey results, especially in *Collaborative Writing with AI Tools and Voice* category, where students' scores increased from pre- to post-module (Pre: *M* = 23.29, *SD* = 4.77; Post: *M* = 31.14, *SD* = 3.68; *Mdn* = 23 [pre], 33 [post]), and all seven students reported post-module gains (e.g., St #4: 20 – 35; St #6: 30 – 35; St #1: 23 – 28). These perceived increases reflect students' increased confidence in working with AI as a writing tool and learning to integrate their voice in academic writing.

While most students reported higher post-module scores across categories, one pattern that complicates a simple 'more is better' narrative is that increased critical awareness may also lead some students to evaluate AI's benefits more cautiously. This means that as they learn more about these tools, they may view them to be more limited than having potentials. In Table 2, for example, one student reported a decline in *AI's Impact on Student Success and Access* (St #1: 32 – 27), even though the overall group mean increased (Pre: *M* = 29.71, *SD* = 3.88; Post: *M* = 35.00, *SD* = 5.83; *Mdn* = 32 [pre], 40 [post]). This shift is compatible with the researcher's classroom observations that students were increasingly interacting with AI outputs more actively rather than accepting them at face value. As the researcher's journal notes: "They were not only writing more thoughtful prompts, but also engaging in prompt chaining – a sign that they were thinking critically and interacting with AI outputs rather than passively accepting them." In other words, as students became more discerning about AI's limitations, some may have simultaneously become less confident that AI is uniformly beneficial for success and access in all academic tasks.

While the researcher observed strong classroom engagement, particularly during hands-on, exploratory, and discussion-based activities during the module, the course instructor, who was able to observe the students across the semester, noted a general lack of engagement from the students with class content throughout the course, which could be a reason for their continuing overreliance on AI tools over the semester as observed by the course instructor. This tension between productive in-class engagement during the piloted instructional module and uneven transfer beyond the module (as observed by the course instructor) also surfaced in Lesson 4, whose goal was "to help students practice everything they had learned in the previous three lessons – voice, AI capabilities and limitations, and effective prompting – by applying those concepts to real-world [academic] writing scenarios." The researcher noted that students' ability to engage with these scenarios seemed to depend also partly on prior writing preparation:

> "One challenge I encountered was with Scenario 8… many students hadn't yet taken [First-Year Composition (FYC) course]. Since students are required to take the FYS and FYC in different semesters, the majority of the class didn't have the foundational knowledge to fully understand the scenario." (Researcher Reflective Journal)

According to her, students who had completed FYC were able to approach the scenario with more confidence and attention to nuance, while others struggled to apply the concepts of voice, revision, quote interpretation in source-based writing, and audience awareness, and in general, engaging with AI in this specific writing task. This suggests that critical, learning-centered AI use in academic writing may depend not only on familiarity with AI tools, but also on students' foundational knowledge of academic writing conventions and revision practices, which also helps explain why post-module perceptions (especially in *AI's Impact on Success and Access*) may become more variable across students.

Finally, the researcher's in-class observations of students' active engagement revising AI outputs with voice and purpose, and the higher post-module survey scores collected right after the instructional module, contrast with the course instructor's broader concerns about student engagement and overreliance on AI across the semester. This pattern suggests that AI literacy activities may be most effective when embedded repeatedly across assignments and genres, rather than implemented as a single module, so that students have multiple opportunities to apply voice strategies, critical evaluation, and purposeful prompting in varied academic writing tasks and across different genres.

Discussion

This pilot study examined how a short instructional module shaped first-year students' perceived AI literacy in academic writing, triangulating (a) pre/post Perceived AI Literacy Survey scores, (b) the researcher's reflective journal, and (c) the course instructor's end-of-semester interview. Across all four survey categories, mean scores increased from pre- to post-module, with the largest gains in *Collaborative Writing with AI Tools and Voice* (Pre $M = 23.29$, $SD = 4.77$; Post $M = 31.14$, $SD = 3.68$) and notable gains in *Use and Application of AI Tools in Academic Contexts* (Pre $M = 25.14$, $SD = 2.10$; Post $M = 31.29$, $SD = 4.37$). These patterns may suggest that the module contributed to shifts in students' confidence in working with AI as a writing partner while maintaining authorship, voice, and revision control, rather than simply increasing enthusiasm for AI. As such, students' strongest perceived growth clustered around *Use, Apply* and *Evaluate* constructs of AI literacy frameworks (Almatrafi, et. al., 2024). This outcome aligns with broader AI literacy scholarship that defines AI literacy as competencies for interacting with and reasoning about AI intentionally, rather than merely knowing about it, scaffolded through task-based, learning-centered and reflective activities that target the development of these skills (Almatrafi et al., 2024; Long & Magerko, 2020; Mills et. al., 2024; Ng, et al, 2021).

The main focus of our instructional module was to help students position AI writing support as collaborative and learn how to foreground voice and revision strategies as they interact with generative AI tools in academic writing contexts. In particular, students' post-module gains in *Collaborative Writing with AI Tools and Voice* category align with long-standing writing scholarship that treats voice as a rhetorical and identity-laden dimension of academic writing (Elbow, 2007; Hyland, 2002; Ivanič, 1998). When students practiced revising AI-generated text to better reflect their stance, tone, and specificity, they were not only learning how to use a tool, but also practicing how to *reclaim ownership* of meaning and argument, which also resonates with AI literacy frameworks that centers human judgment and agency when using AI-enabled tools (Mills et al., 2024). Our findings and observations also complement Tate et al. (2025), who argue that generative AI can function as a *productive writing partner* only when instruction is designed to preserve students' authorial agency. In their design-based implementation of an AI-supported writing approach, they emphasize that students learn most when they engage in iterative prompting and reflection by treating AI output as a revisable resource and making intentional decisions about how (and whether) to incorporate it into their drafts.

Prior instructional interventions have suggested that AI literacy develops most robustly when students engage in task-based, hands-on, learning-centered activities that position AI as something to *work with* and *think against*, rather than something to passively accept. Systematic reviews of K–12 AI literacy work show that learning experiences commonly emphasize recognizing AI artifacts, building conceptual understanding of how AI works, and critically evaluating AI's implications through contextualized, rather than decontextualized, activities (Casal-Otero et al., 2023). Similarly, empirical examples, such as the project-based experiences like the DAILy curriculum (Lee et al., 2021) and video and graphic design projects (Akman, 2025), gamified simulations such as the AI Audit card game (Ali et al., 2023), and goal-based scenario learning designs (Tseng et al., 2024), further demonstrate that learners need to be given

structured design-based opportunities to test AI's capabilities and limits while also foregrounding ethical and societal considerations. In higher education writing contexts, similar active-learning approaches have been shown to shift students toward more skeptical, metacognitive engagement with AI outputs (van Niekerk et al., 2025). Our findings align with this intervention literature: the task-based, discussion-driven, scaffolded, and reflective activities in our module appeared to support students' developing agency (particularly around voice and revision) and more critical engagement with AI's affordances and constraints, suggesting that 'learning-by-doing' designs may be important for moving students from passive AI consumption toward strategic, learning-centered use.

Across our cross-analysis of survey data, the researcher's journal, and the course instructor's end-of-semester interview, an important finding was a transfer gap, meaning that the observed improvements in students' ethical reasoning and attention to personal voice in academic writing during the activities in the instructional module did not carry over to later, independent course assignments, as reported in the course instructor's end-of-semester interview. This pattern is consistent with the view in AI literacy frameworks that ethical engagement is not a one-time achievement but a context-sensitive practice shaped by task design, incentives, and what students perceive to 'count' as learning (Almatrafi et al., 2024; Mills et al., 2024). At the same time, the course instructor's later observation about the students' tendency to rely on AI-generated interviews rather than conducting authentic interviews themselves illustrates a familiar challenge in writing pedagogy: strategies developed in scaffolded learning contexts may not generalize automatically to new tasks unless instruction explicitly supports transfer across contexts (Wardle, 2007). This pattern may reflect a broader challenge identified in learning sciences research: knowledge developed in structured instructional contexts does not always transfer to new tasks unless learners recognize when and how previously learned concepts apply.

This disconnection reflects what Bransford, Brown, and Cocking (2000) describe as a *failure of learning transfer*, or the ability to apply knowledge and skills to new contexts. As they note, "students often fail to transfer knowledge because they do not recognize when it is relevant" (Bransford et al., 2000, p. 53). Similarly, from a situated learning perspective, knowledge and practices are often tied to the specific contexts in which they are learned (Lave & Wenger, 1991), meaning that students may not automatically apply ethical AI practices developed in guided classroom activities to different assignments unless those practices are repeatedly embedded across learning environments. In our context, students may have articulated ethical principles within guided scenario-based discussions but did not recognize that those same principles should guide decision-making in unfamiliar, less structured, and higher-stakes writing assignments later in the semester. This also connects to Ng et al.'s (2022) argument that meaningful AI literacy requires a practical understanding of AI's potentials, challenges, and limitations, including knowing "when and how to use it ethically" (p. 122), which must be internalized to guide authentic behavior beyond instructional discussions. Bransford et al. (2000) further emphasize that transfer depends heavily on the conditions of initial learning, noting that "organizing information into a conceptual framework allows for greater transfer" (p. 17). While our instructional module's guided discussions and activities likely supported such conceptual organization in the moment, those frameworks may not have been cued or activated later without continued scaffolding, reinforcement, and task designs that make ethical decision-making unavoidable and visible.

Finally, transfer of learning also requires sufficient time and repeated opportunities to apply learning across contexts, as Bransford et al. (2000) remind us that "the learning of complex content cannot be rushed" (p. 58). All in all, it can be conclude that the module likely provided an important starting point for ethical awareness, but sustained transfer may require recurring, assignment-embedded opportunities that repeatedly cue ethical and learning-centered AI use across genres and tasks (Almatrafi et al., 2024; Mills et al., 2024).

Although our study did not directly measure students' prior writing preparation or enrollment in First-Year Composition (FYC), our triangulated observations suggest that foundational academic writing knowledge may influence how students evaluate and engage with generative AI. This interpretation aligns with Kim, Ham, and Lee's (2024) findings that students with stronger domain-relevant skills and more positive orientations toward AI demonstrated more reflective, iterative, and metacognitively regulated interactions with the tool. Similarly, our findings support the view that critical AI literacy in academic writing is not simply 'tool literacy,' but a situated practice requiring familiarity with genre expectations, revision strategies, and rhetorical decision-making about evidence, interpretation, and audience (Wang & Wang, 2025). When students have stronger foundations in academic writing, they may be better able to discern when AI outputs support learning rather than replacing it. In contrast, students with less confidence or experience in these areas may be more likely to rely heavily on AI-generated text without critical evaluation. While this was not a measured outcome in our study, the parallel with Kim et al.'s high-

skill groups suggests that prior domain knowledge may provide an important foundation for ethical, autonomous, and critically engaged AI use. It also helps explain why perceived gains may be uneven across the students in our study and why some struggled with more advanced source-based and revision-intensive scenarios. These concerns also align with research showing that while many students use ChatGPT for writing-related tasks, they often remain uncertain about appropriate boundaries and integrity implications (Črček & Patekar, 2023; Gruenhagen et al., 2024).

Taken together, this study contributes to emerging research on generative AI in higher education in three ways. First, it provides an instructional design example demonstrating how AI literacy instruction can be integrated into first-year seminar and writing contexts through activities that combine ethical AI use, prompting strategies, and explicit attention to student voice. Second, the study contributes to writing pedagogy by framing generative AI not simply as a tool to regulate or prohibit, but as a collaborative writing partner whose outputs require critical revision and critical decision-making. Third, although exploratory in scope, the study offers early empirical insight into how first-year students perceive and negotiate the use of generative AI in academic writing tasks, including the challenges of transferring ethical AI practices beyond structured instructional contexts.

This study faced several limitations that should be acknowledged. First, the student sample in our study only represents a portion of the students in one class at one university. Because of the small sample in this study, the results of this study cannot be generalized to other populations, and nor was this intended to. The findings should be taken as descriptive, rather than conclusive. Additionally, the study relied on self-reported perceptions of AI literacy rather than direct assessments of students' AI-related writing practices or performance. While perception-based measures provide insight into students' awareness and confidence, future studies should incorporate behavioral or performance-based assessments to more directly examine AI literacy development. In addition, subsequent studies should consider implementing the similar modules with larger student populations, preferably representing more than one class.

Additionally, despite being offered the opportunity, no students volunteered for follow-up interviews after the module implementation. This limited our ability to gather in-depth qualitative insights into students' perceptions and experiences. Incorporating open-ended questions within the survey instrument in future iterations of the study may help capture richer qualitative data while accommodating students' time constraints and willingness to participate. In addition, adding reflection-based assignments in the instructional module through which students reflect on their learning processes would support deeper metacognitive engagement and offer a valuable window into students' evolving understanding.

Moreover, the two-week timeframe allocated for the module delivery, which adhered to ethical considerations required by the IRB approval for this study, was relatively short, especially considering that the the researcher was only the module instructor but not the course instructor. This created challenges in terms of continuity, relationship-building, and sustained engagement with students. It is suggested that having the course instructor also serve as the module facilitator in the future would enhance integration of the content into the course and allow closer observation of student progress throughout the semester.

Additionally, this alignment would create ongoing opportunities to implement and practice the module content through regular class assignments, reinforcing key concepts of AI literacy in more authentic and scaffolded ways.

Finally, the timing of the collection of the post-survey data presents limitations. While the immediate post-module survey responses indicated positive shifts in students' understanding, these results may not reflect the longer-term effects of the intervention. The course instructor, interviewed only at the end of the semester – a time typically marked by fatigue and increased workload – expressed more tempered views about the module's impact. A more comprehensive understanding could have been achieved by conducting interviews at multiple points, such as immediately after the module's delivery and again at semester's end.

Implications

The findings of this study offer important implications for the design and implementation of AI literacy instruction in Higher Education settings. First of all, it is important that students receive focused instruction on how to recognize, use and evaluate AI, such as the instructional module designed for this study, to develop their AI literacy skills. Such focused instruction should involve hands-on tasks, allow student reflection, and provide space for discussion, all of which encourage active engagement with generative AI tools. Since AI use and ethical evaluation will depend on the

particular course or task, it may also be beneficial for HE instructors to make such focused instruction part of their specific courses to guide students.

As AI literacy lies at the intersections of traditional literacy and digital literacy, the findings suggest that students' foundational academic writing knowledge reinforces the development of AI literacy. If students do not have knowledge and experience in academic writing, or if they do not have awareness of their own personal voice in academic writing, their critical evaluation of the AI-generated academic writing may be limited. Therefore, it is recommended that integrating AI literacy instruction with, or sequencing it after, writing instruction that builds rhetorical awareness. For example, institutions could consider embedding it within FYC courses to ensure students have the rhetorical tools to use AI as a support rather than a substitute for learning.

Finally, this study reinforces the need for reassessing how student work is evaluated in the age of generative AI. While students may demonstrate critical awareness and thoughtful engagement in class, such behaviors are not always visible in current final assignments and traditional assessment methods. Instructors may need to adopt more process-based and reflection assessment tools – such as annotated drafts, process journals, or AI-use reflections – to gain a fuller picture of how students are using these tools. These practices not only uphold academic integrity but also encourage responsible, intentional use of AI.

Conclusion

This pilot study highlights both the promise and the limitations of short-term AI literacy interventions in first-year seminar classrooms. The instructional module demonstrated potential to increase students' knowledge and ability to use and apply AI tools in academic contexts, their sense of authorial voice when engaging with generative AI tools, and their critical evaluation skills. Survey results and classroom observations revealed early signs of critical literacy, particularly when students were guided through scaffolded, hands-on, reflective activities. However, these gains did not consistently transfer to independent academic tasks, suggesting that deeper, more durable understanding requires more than one-time instructional intervention.To foster truly reflective, voice-driven, and context-aware AI use, AI literacy focused instruction needs to be integrated across the curriculum – especially in writing-intensive courses – through repeated, varied opportunities for application.

References

Agyare, B., Asare, J., Kraishan, A., Nkrumah, I., & Adjekum, D. K. (2025). A cross-national assessment of artificial intelligence (AI) Chatbot user perceptions in collegiate physics education. *Computers and Education: Artificial Intelligence, 8,* 100365. https://doi.org/10.1016/j.caeai.2025.100365

Akman, E. (2025). The impact of AI-based visual designs on students' AI literacy and attitudes toward AI. *Interactive Learning Environments, 33*(8), 1–19. https://doi.org/10.1080/10494820.2025.2530630

Ali, S., Kumar, V., & Breazeal, C. (2023). AI Audit: A card game to reflect on everyday AI systems. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 37, No. 13, pp. 15981–15989). AAAI Press. https://doi.org/10.1609/aaai.v37i13.26897

Almatrafi, O., Johri, A., & Lee, J. (2024). A systematic review of AI literacy conceptualization, constructs, and implementation and assessment efforts (2019–2023). *Computers and Education Open, 5,* 100135. https://doi.org/10.1016/j.caeo.2024.100135

Amrullah, M. I., Firmansyah, N., Nabilah, A., Hendriwanto, & Muntiningsih, R. (2024). ChatGPT in language education: Meta-analysis. *Journal of World Englishes and Educational Practices, 6*(2), 18–25. https://doi.org/10.32996/jweep.2024.6.2.2

Ateriya, N., Sonwani, N. S., Thakur, K. S., Kumar, A., & Verma, S. K. (2025). Exploring the ethical landscape of AI in academic writing. *Egyptian Journal of Forensic Sciences, 15*(1), 36. https://doi.org/10.1186/s41935-025-00453-1

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school* (Expanded ed.). National Academy Press.

Bouzar, A., El Idrissi, K., & Ghourdou, T. (2024). *ChatGPT and academic writing self-efficacy: Unveiling correlations and technological dependency among postgraduate students*. Arab World English Journal (AWEJ), Special Issue on ChatGPT, 225–236. https://doi.org/10.24093/awej/ChatGPT.15

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Casal-Otero, L., Rodríguez-García, A. M., Gago-Rodríguez, S., Vázquez-Cano, E., & López-Meneses, E. (2023). AI literacy in K–12 education: A systematic review of concepts, purposes, methods, and effectiveness of

interventions (2017–2023). *International Journal of STEM Education, 10*(1), 36. https://doi.org/10.1186/s40594-023-00418-7

Črček, N., & Patekar, J. (2023). Writing with AI: University students' use of ChatGPT. *Journal of Language and Education, 9*(4), 128–138. https://doi.org/10.17323/jle.2023.17379

Cummings, R. E., Monroe, S. M., & Watkins, M. (2024). *Generative AI in first-year writing: An early analysis of affordances, limitations, and a framework for the future.* Computers and Composition, 71, Article 102827. https://doi.org/10.1016/j.compcom.2024.102827

Dai, Y., Liu, A., Qin, J., Guo, Y., Jong, M. S.-Y., Chai, C.-S., & Lin, Z. (2023). Collaborative construction of artificial intelligence curriculum in primary schools. *Journal of Engineering Education, 112*(1), 23–42. https://doi.org/10.1002/jee.20503

Das, S.R. & J.V., M. (2024). Perceptions of higher education students towards ChatGPT usage. *International Journal of Technology in Education (IJTE), 7*(1), 86-106. https://doi.org/10.46328/ijte.583

Fauzi, F., Tuhuteru, L., Sampe, F., Ausat, A. M. A., & Hatta, H. R. (2023). Analysing the role of ChatGPT in improving student productivity in higher education. *Journal on Education, 5*(4), 14886–14891.

Gu, X., & Ericson, B. J. (2025, August). Ai literacy in K-12 and higher education in the wake of generative AI: An integrative review. In *Proceedings of the 2025 ACM Conference on International Computing Education Research V. 1* (pp. 125–140).

Gupta, S., Jaiswal, A., Paramasivam, A., & Kotecha, J. (2022). Academic writing challenges and supports: Perspectives of international doctoral students and their supervisors. *Frontiers in Education, 7,* 891534. https://doi.org/10.3389/feduc.2022.891534

Gruenhagen, J. H., Sinclair, P. M., Carroll, J.-A., Baker, P. R. A., Wilson, A., & Demant, D. (2024). *The rapid rise of generative AI and its implications for academic integrity: Students' perceptions and use of chatbots for assistance with assessments*. *Computers and Education: Artificial Intelligence, 7*, 100273. https://doi.org/10.1016/j.caeai.2024.100273

Healey, M., Matthews, K. E., & Cook-Sather, A. (2020). *Writing about learning and teaching in higher education: Creating and contributing to scholarly conversations across a range of genres* (pp. 275–283). Elon University Center for Engaged Learning. https://doi.org/10.36284/celelon.oa3

Holland, A., & Ciachir, C. (2024). A qualitative study of students' lived experiences and perceptions of using ChatGPT: immediacy, equity and integrity. *Interactive Learning Environments,* 1–12.

Kim, J., Ham, Y., & Lee, S. S. (2024). Differences in student-AI interaction process on a drawing task: Focusing on students' attitude towards AI and the level of drawing skills. *Australasian Journal of Educational Technology, 40*(1), 19–41. https://doi.org/10.14742/ajet.8859

Kim, J., Yu, S., Detrick, R., & Li, N. (2024). Exploring students' perspectives on generative AI-assisted academic writing. *Education and Information Technologies, 30,* 1265–1300. https://doi.org/10.1007/s10639-024-12878-7

Lee, I., Ali, S., Zhang, H., DiPaola, D., & Breazeal, C. (2021). Developing middle school students' AI literacy. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education (SIGCSE '21)* (pp. 191-197). ACM. https://doi.org/10.1145/3408877.3432513

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press.

Long, D. & Magerko, B. (2020). What is AI literacy? Competencies and Design Considerations. *Proceedings of the 2020 CHI Conference on Factors in Computing Systems,* 1-16. https://doi.org/10.1145/3313831.3376727

Mills, K., Ruiz, P., Lee, K., Coenraad, M., Fusco, J., Roschelle, J., & Weisgrau, J. (2024, May). AI Literacy: A Framework to Understand, Evaluate, and Use Emerging Technology. https://doi.org/10.51388/20.500.12265/218

Mollick, E. R., & Mollick, L. (2023). *Assigning AI: Seven approaches for students, with prompts* (The Wharton School Research Paper). SSRN. https://doi.org/10.2139/ssrn.4475995

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

Ng, D. T. K., Leung, J. K. L., Su, M. J., Yim, I. H. Y., Qiao, M. S., & Chu, S. K. W. (2022). *AI literacy in K–16 classrooms.* Springer International Publishing. https://doi.org/10.1007/978-3-031-18880-0

Onal, S., Kulavuz-Onal, D., & Childers, M. (2025). Patterns of ChatGPT usage and perceived benefits on academic performance across disciplines: Insights from a survey of higher education students in the United States. *Journal of Educational Technology Systems, 54*(3), 1–33. https://doi.org/10.1177/00472395251341214

OpenAI. (2025). *ChatGPT (May 14 version) [Large language model].* https://chat.openai.com/https://chat.openai.com/

Elbow, P. (2007). *Voice in writing again: Embracing contraries*. *College English, 70*(2), 168–188.

Promma, W., Imjai, N., Usman, B., & Aujirapongpan, S. (2025). The influence of AI literacy on complex problem-solving skills through systematic thinking skills and intuition thinking skills: An empirical study in Thai Gen Z accounting students. *Computers and Education: Artificial Intelligence, 8,* 100382. https://doi.org/10.1016/j.caeai.2025.100382

Saldana, J. (2021). *The coding manual for qualitative researchers* (4th ed.). Thousand Oaks, CA: SAGE.

Tarchi, C., Zappoli, A., Casado Ledesma, L., & Brante, E. W. (2024). The use of ChatGPT in source-based writing tasks. *International Journal of Artificial Intelligence in Education.* https://doi.org/10.1007/s40593-024-00413-1

Tate, T. P., Harnick-Shapiro, B., Ritchie, D. R., Tseng, W., Dennin, M., & Warschauer, M. (2025). *Incorporating generative AI into a writing-intensive undergraduate course without off-loading learning*. Discover Computing, 28, Article 72. https://doi.org/10.1007/s10791-025-09563-9

Tseng, Y.-J.; Yadav, G.; Hou, X.; Wu, M.; Chou, Y.-S.; Chen, C. C.; Wu, C.-C.; Chen, S.-G.; Lin, Y.-J.; Liao, G.; et al. 2024b. ActiveAI: The Effectiveness of an Interactive Tutoring System in Developing K-12 AI Literacy. In *European Conference on Technology Enhanced Learning*, 452-467. Springer.

Van Niekerk, J., Delport, P. M. J., & Sutherland, I. (2025). Addressing the use of generative AI in academic writing. *Computers and Education: Artificial Intelligence, 8,* 100342. https://doi.org/10.1016/j.caeai.2024.100342

Vengadasalam, S. S. (2020). *Transformative pedagogy and student voice: Using S.E.A. principles in teaching academic writing*. *Journal of Effective Teaching in Higher Education, 3*(2), 12–27.

Wang, C. (2024). *Exploring students' generative AI-assisted writing processes: Perceptions and experiences from native and nonnative English speakers.* Technology, Knowledge and Learning. https://doi.org/10.1007/s10758-024-09744-3

Wang, C., & Wang, Z. (2025). Investigating L2 writers' critical AI literacy in AI-assisted writing: An APSE model. *Journal of Second Language Writing, 67*, 101187. https://doi.org/10.1016/j.jslw.2025.101187

Wardle, E. (2007). Understanding "transfer" from FYC: Preliminary results of a longitudinal study. *WPA: Writing Program Administration, 31*(1–2), 65–85.

Wiese, L. J., Patil, I., Schiff, D. S., & Magana, A. J. (2025). AI ethics education: A systematic literature review. *Computers and Education: Artificial Intelligence, 8,* 100405. https://doi.org/10.1016/j.caeai.2025.100405

Yang, H., Huang, R. H., & Liu, D. (2025). Adapting teaching and learning with existing generative AI tools: A teacher–student co-evolution approach. *Computers and Education: Artificial Intelligence, 8,* 100366. https://doi.org/10.1016/j.caeai.2024.100366

Zhao, X., Cox, A., & Cai, L. (2024). ChatGPT and the digitisation of writing. *Humanities and Social Sciences Communications, 11,* Article 482. https://doi.org/10.1057/s41599-024-02904-x

Effect of an Inclusive Education Course Delivered Through a Flipped Learning Approach on Pre-Service Teachers' Self-Efficacy: A Randomized Controlled Trial

Yogaranee SAKTHIVEL
Department of Educational Psychology, Faculty of Education, University of Colombo, Sri Lanka
yogaranee@edpsy.cmb.ac.lk
ORCID:0009-0000-7803-3206

ABSTRACT
The discrimination and exclusion of children with special educational needs from mainstream classrooms can be attributed to inadequate preparation of pre-service teachers for inclusive teaching. Teachers' self-efficacy, which refers to beliefs about their ability to implement inclusive strategies, is widely recognized as a crucial factor in preparing pre-service teachers to perform confidently and effectively in mainstream classrooms. This article, part of a broader study exploring the mediating role of pre-service teachers' beliefs about inclusive education in the relationship between an inclusive education course intervention- delivered through a flipped learning approach- and teachers' self-efficacy, reports only the direct effects. The sample consisted of 240 Tamil-speaking pre-service teachers randomly selected from initial teacher education institutions across Sri Lanka. A randomized controlled trial with pre- and post-tests, employing an experimental design with a control group, was conducted. Data were collected using the culturally adapted Tamil version of the Teacher Efficacy for Inclusive Practices scale and analyzed through MANCOVA in IBM SPSS and PLS-SEM in SmartPLS4. The results revealed the differential impact of the intervention on the three domains of teacher efficacy: inclusive instruction, behaviour management, and collaboration, with the intervention significantly predicting these efficacy dimensions. These findings contribute to initial teacher education programs by providing structured, institution-level, course-based exposure to inclusive education through practical approaches, thereby better preparing pre-service teachers for inclusive classrooms.
Keywords: Inclusive education course, Inclusive practices, Pre-service teachers, Randomized controlled trial, Teacher efficacy.

INTRODUCTION
The global movement towards inclusive education (IE) has created significant challenges for pre-service teachers (PSTs), especially as increasing numbers of students with special educational needs (SEN) are integrated into mainstream classrooms. PSTs are expected to address long-standing discriminatory attitudes and exclusionary practices that continue to persist within inclusive settings (O'Neill, 2015; UNESCO, 2008). The adequate preparation of PSTs has become imperative, requiring the development of IE-related knowledge, practical skills, and positive professional dispositions, a concern that remains global in scope (UNICEF, 2013). Evidence from reviews of initial teacher education (ITE) programs across the Asia-Pacific region further indicates that many systems continue to face substantial challenges in equipping graduates with the essential competencies needed to manage learner diversity effectively (UNICEF, 2013). This gap can largely be attributed to ITE programs that have historically provided limited exposure to IE alongside university-based coursework, thereby leaving PSTs insufficiently prepared for the practical challenges of their future classrooms (Alazemi & Larkins, 2025; Avramidis et al., 2000; Carroll et al., 2003; Lancaster & Bain, 2010; Lucena-Rodríguez et al., 2025).

Like many developing countries, Sri Lanka has demonstrated strong policy commitments to IE. However, a considerable gap persists between policy mandates and actual practice in ITE. Despite significant efforts by ITE institutions to prepare PSTs to implement IE effectively, many of them report feeling overwhelmed by student diversity and concerned about meeting the varied needs of students with SEN. They attribute these concerns to ITE programs, perceiving them as overly theoretical and insufficiently connected to their future classroom requirements (Hettiarachchi & Das, 2014; Yogaranee, 2024, 2025). A recent nationwide survey by Peries et al. (2021) involving 705 primary teachers revealed that although most held positive attitudes towards inclusion, many did not feel prepared to identify and support students with dyslexia, underscoring the need to enhance ITE with more structured, practical experiences with students with SEN and IE. To this end, this study employs a randomised controlled trial (RCT) to examine how an IE course, aligned with the Sri Lanka Qualifications Framework and delivered via a flipped learning approach, can improve PSTs' confidence in applying inclusive methods, particularly in teaching, behaviour management, and collaboration.

LITERATURE REVIEW

Teachers' sense of efficacy (TSE), introduced in Bandura's (1997) social cognitive theory, refers to teachers' belief in their ability to achieve teaching goals. Bandura suggests that TSE is influenced by four primary sources of information: mastery experiences, physiological and emotional states, vicarious experiences, and verbal persuasion. ITE programs are expected to provide some or all of these experiences to help develop TSE. Bandura highlighted that mastery experiences are the most crucial; therefore, measures of self-efficacy should be sensitive to how such experiences during ITE affect PSTs' attitudes and confidence in supporting students with SEN.

Developing TSE for inclusive practices among teachers is a crucial endeavour, supported by evidence linking it to teachers' retention, students' academic adjustment, and teachers' professional fulfilment (Bandura, 1997; Zee & Koomen, 2016). TSE is also negatively associated with burnout, a state characterized by emotional exhaustion, depersonalization, and reduced personal achievement, which can cause teachers to leave the profession prematurely (Brouwers & Tomic, 2000). High-efficacious teachers tend to demonstrate greater resilience when facing classroom challenges. Conversely, those with lower TSE are more likely to adopt punitive, authoritarian classroom management strategies that limit student autonomy and engagement.

It is widely reported that the early years of pre-service training are the most effective time to influence PSTs' beliefs and confidence towards IE (Woolfolk-Hoy & Spero, 2005). Experts argue that this period is ideal for fostering positive attitudes and developing TSE through high-quality training (Lambe & Bones, 2007). To prepare PSTs as inclusive practitioners, ITE programs must equip them with sufficient inclusive knowledge and skills to challenge their assumptions about fairness and equity. Beyond just knowledge, teachers need to develop positive values, supportive ideals, and a strong sense of responsibility for educating all children, regardless of their diverse needs (Forlin, 2010). Without this foundation, teachers are unlikely to commit to inclusive classrooms fully, highlighting the importance of reforms that realign ITE with inclusive goals.

The shift towards fully inclusive school systems requires that ITE institutions revise their curricula to incorporate IE courses, either as standalone modules or embedded content, to prepare PSTs better (Forlin, 2010; Kurniawati et al., 2014). However, many curricula failed to fulfil their requirements, instead relying heavily on a single compulsory IE course, which offered limited opportunities for extended practicum experiences, electives, or in-depth study (Carroll et al., 2003). Consequently, many new teachers report feeling anxious about their ability to teach students with diverse learning needs and attribute these concerns to inadequate preparation for inclusive practice (Forlin & Chambers, 2011; Hemmings & Weaven, 2005; Lambe & Bones, 2007; Sharma & Nuttall, 2016; Yogaranee, 2024).

TSE in inclusive practices has been the central area of most international studies examining how courses on IE or special education influence PSTs' TSE. They consistently demonstrated that participation in a special or IE course positively impacts attitudes and TSE of PSTs (Casarez, 2013; Can, 2015; Forlin et al., 2014; Lancaster & Bain, 2007, 2010; Loreman et al., 2013; O'Neill, 2015; Sharma, 2012; Sharma et al., 2008; Sharma et al., 2015; Sharma & Nuttal, 2016; Sharma & Sokal, 2015; Woodcock et al., 2012; Zundans-Fraser & Lancaster, 2012).

Lancaster and Bain's (2007) study examined the effectiveness of three different design conditions: two involved direct experience with students with SEN, and one followed a university 'subject-only' approach. Significant improvements in TSE were seen across all groups, with no notable differences between the approaches. In their follow-up study in 2010, they compared two versions of a 13-week required IE course, one including a field placement and the other based on complex adaptive systems. Both formats led to significant gains in TSE among PSTs for inclusive teaching, with no differences between them. This indicates that well-structured alternative courses can be as effective as practicum-based experiences.

On the contrary, Nketsia and Saloviita (2013), who conducted a study on PSTs' views on IE in Ghana, found that although nearly all participants had been exposed to the concept of inclusion during their studies, only about one-third reported feeling highly or somewhat prepared to teach learners with diverse abilities. This finding highlights a gap between exposure to inclusion as a theoretical concept and the development of self-efficacy to implement IE.

Despite the growing emphasis on inclusion, research consistently indicates that PSTs often feel inadequately prepared to teach students with SEN, reporting low levels of TSE for inclusive practices (Forlin & Chambers, 2011). It has also been shown that ITE programs are increasingly including modules on IE, which tend to be theoretical and offer limited opportunities for practical application, thus restricting their impact on the development of TSE (Sharma et al., 2013). Furthermore, there is insufficient experimental evidence to determine whether structured IE courses can effectively enhance PSTs' TSE.

The current study identified several gaps, including conceptual, methodological, and statistical issues within the reviewed literature. Firstly, the targeted intervention of the IE course aimed at developing TSE is missing in the Sri Lankan context, particularly for Tamil-speaking PSTs. Secondly, empirical evidence does not demonstrate the inclusion of experimental interventions, especially in an RCT with a pre- and post-test controlled-group design. Thirdly, research examining the effect of the IE course on PSTs' TSE has consistently employed a pre- and post-test design, using various inferential statistics, such as paired-sample t-tests, one-way ANOVAs, one-way MANCOVAs, and repeated-measures ANOVAs, without employing structural equation modelling (SEM) techniques. Lastly, the study highlights the design of IE, which is widely emphasised in most studies, involving both faculty-led course-based exposure and a practicum that facilitates mastery of the experiences gained through the targeted intervention of the IE courses. However, some studies focus solely on the design of the intervention, using course-based exposure alone, which has been reported to be sufficient for improving TSE in inclusive practices.

This study addresses these gaps by investigating how effective the targeted experimental, course-based intervention of the IE course is in fostering TSE for inclusive practices among PSTs, evaluated across three areas: efficacy for inclusive instruction (EII), efficacy for managing behaviour (EMB), and efficacy for collaboration (EC). The following research questions were formulated: (1) Does the IE course lead to higher post-test EII scores in the experimental (EXP) group compared to the control (CON) group, after adjusting for baseline EII scores? Does the intervention also significantly predict post-test EII scores? (2) Does the IE course result in higher post-test EMB scores in the EXP group compared to the CON group, after adjusting for baseline EMB scores? Does the intervention also significantly predict post-test EMB scores? (3) Does the IE course generate higher post-test EC scores in the EXP group compared to the CON group, after adjusting for baseline EC scores? Does the intervention also significantly predict post-test EC scores? It was hypothesised that the IE course intervention would significantly predict improvements in PSTs' overall TSE for inclusive practices, such as EII, EMB, and EC, with the EXP group showing higher post-test mean scores than the CON group for the total Teacher Efficacy for Inclusive Practices (TEIP) scale and its three dimensions—EII, EMB, and EC—after adjusting for the corresponding pre-test scores.

The findings of this study are important in showing how targeted interventions in the IE course, especially in their design and delivery, can improve PSTs' TSE for inclusive practices. Although the IE course is an institutional-level, course-based programme, implementing it in simulated environments without enough real classroom experience notably helped boost teaching confidence, including instructional techniques, behaviour management, and collaboration among PSTs. The implication is that mastery experiences, a key source of efficacy information, can be recreated through well-structured microteaching and other simulated activities; when the course is delivered effectively, these simulated mastery experiences can enhance PSTs' TSE (Bandura, 1997; Pendergast et al., 2011). These findings show the potential of carefully designed, simulation-based IE coursework to serve as an effective alternative to direct classroom experience in preparing PSTs for inclusive teaching.

METHODS

Research Design

This longitudinal study, part of a broader project examining the mediating role of PSTs' beliefs about IE in the relationship between the IE course intervention, delivered through a flipped learning approach, and TSE, specifically investigated the direct effect of the course intervention on TSE. An RCT design with a pre- and post-test control group was selected, following the Consolidated Standards of Reporting Trials (CONSORT) 2010 guidelines (Moher et al., 2010). The RCT is considered the most rigorous method for evaluating causal effects of educational interventions, with the EXP versus CON group serving as a categorical independent variable and the TSE as the outcome variable (Moher et al., 2010; Schulz & Grimes, 2002).

Participants

The target population includes all Tamil-speaking PSTs enrolled at ITE institutions across Sri Lanka. Three institutions were randomly chosen to create a sampling frame of cohorts from the 2023/25 batches. A total of 240 teacher candidates were randomly selected from this frame as the sample. Although the relatively small proportion of male participants (24%) in the sample might raise concerns about bias, this distribution reflects the national teacher population, in which females are predominant.

Measures

Data were gathered using the Tamil version of the adapted 18-item TEIP scale (TEIP-TM), with permission from the first author (Sharma et al., 2012). The original TEIP scale assesses TSE across three domains: EII, EMB, and EC, each comprising six items, rated on a 6-point Likert-type scale, from "Strongly Disagree" to "Strongly Agree",

with total scores ranging from 18 to 108, where higher scores indicate greater efficacy. Although the TEIP scale developers reported satisfactory internal consistency for the three subscales and the total scale, with Cronbach's alpha values of .93 for EII, .85 for EC, .85 for EMB, and .89 for the overall scale, they did not include comprehensive analyses for factorial structure and construct validation. Therefore, although the 18-item TEIP demonstrates acceptable reliability, further validation is required to establish its measurement properties fully. Nonetheless, the extensive international use of the TEIP scale to assess perceived TSE for teaching in inclusive classrooms supports its relevance and practical utility, rendering it a suitable and well-established instrument for evaluating PSTs' TSE in this study.

The cross-cultural adaptation and validation of the 18-item TEIP-TM scale were undertaken using EFA in IBM SPSS (version 25) and CFA in IBM AMOS (version 23), following established guidelines by Beaton et al. (2000) and the International Test Commission (2016) to assess TSE for inclusive practices among Tamil-speaking PSTs (Sakthivel, in press). EFA, employing principal axis factoring and Direct Oblimin rotation, yielded a three-factor structure for both pre- and post-test scores, aligning with the theoretical factor structure and demonstrating that participants distinctly differentiated among the latent factors of the TEIP-TM: EII, EMB, and EC.

EFA revealed a three-factor structure of the TEIP-TM scale, consistent with the theory, with initial eigenvalues of 7.713 for EII, 2.046 for EMB, and 1.439 for EC. Parallel analysis further supported this three-factor solution. All items on the TEIP-TM scale loaded adequately onto their respective latent constructs, with loadings ranging from .646 to .742 for EII_POST, from .600 to .801 for EMB_POST, and from .592 to .794 for EC_POST, most exceeding the .70 threshold (Hair et al., 2019). Although a few items performed poorly in EFA (EII1_POST, EII3_POST, EMB1_POST, EMB2_POST, and EC1_POST), they were retained for the validation process. EII explained 40.40%, EMB 9.03%, and EC 5.41% of the variance, accounting for 54.84% in total, with EII as the dominant factor. Inter-factor correlations ranged from .454 to .617, indicating related but distinct factors and supporting the use of oblique rotation. All correlations were below .70, indicating no multicollinearity (Field, 2018).

Notably, the previous study adapted the factor structure of the TEIP-TM by including an item from EMB_POST (EMB1 - "I have confidence in my ability to avoid disruptive behaviour in the classroom.") in the EII, due to its cross-loading, while maintaining its dimensionality. This adaptation is both statistically and theoretically justified, as statistically, the primary loading value exceeds the secondary loading by more than 2.0. Theoretically, EMB1_POST is more relevant to include in the EII, as it reflects a preventive management strategy whose function differs from a punishment-oriented, response-based behaviour management strategy. It can be explained that PSTs may have perceived preventive management behaviour as part of inclusive instructional practices rather than as reactive (Sakthivel, 2025; Yogaranee, 2025).

The factorial validity of the three-factor TEIP-TM scale was confirmed using the 7-item EII_POST (including EMB1_POST), the 5-item EMB_POST, and the 6-item EC_POST, performed in IBM AMOS. The model fit indices, including CMIN/df, CFI, TLI, RMSEA, and SRMR, indicated excellent fit to the data. The standardized lambda values were significant, ranging from .615 to .806 for EII_POST, from .629 to .791 for EMB_POST, and from .603 to .793 for EC_POST.

Composite reliability (CR) indices of .877 for EII_POST, .859 for EMB_POST, and .886 for EC_POST all exceeded the recommended threshold of ≥ .70 (Hair et al., 2019; Fornell & Larcker, 1981). Cronbach's alpha values were .880 for EII_POST, .857 for EMB_POST, .884 for EC_POST, and .920 for the total scale, exceeding the ≥ .70 thresholds, indicating strong internal consistency of the scale and subscales (Nunnally & Bernstein, 1994; Hair et al., 2019; Sharma et al., 2012). The average variance extracted (AVE) values, which reflect convergent validity, were .515 for EII_POST, .551 for EMB_POST, and .564 for EC_POST, all exceeding the ≥ .50 threshold, suggesting that each construct explained more than 50% of the variance in its indicators. Furthermore, discriminant validity was confirmed through the heterotrait-monotrait (HTMT) ratio and the Fornell–Larcker (F-L) criterion in the previous study (Fornell & Larcker, 1981; Sakthivel, in press).

As all the factors and the scale met the validity and reliability criteria, it was decided to retain the items with low loadings extracted from the EFA and the adapted factor structure. Overall, previous international studies have consistently confirmed the dimensionality and psychometric soundness of the 18-item TEIP scale. The culturally adapted validation results further aligned with this evidence, supporting the TEIP-TM scale's suitability for assessing TSE for inclusive practices among Tamil-speaking PSTs in Sri Lankan ITE contexts (Sakthivel, in press).

Procedure

Participants were randomly assigned to either the EXP group (N = 120), which received the intervention, or the CON group (N = 120), which followed the standard teacher education courses. Random assignment ensured comparability between groups and minimized selection bias (Torgerson & Torgerson, 2008). Both the EXP and CON groups completed the cross-culturally translated and adapted TEIP-TM scale before (pre-test) and after the intervention (post-test). This design enabled the assessment of both within-group changes over time and between-group differences at post-test, while adjusting for initial TSE levels.

Randomization, allocation concealment, and attrition management

A random sequence was generated for 240 participants, who were randomly assigned to the EXP and CON groups using a computer-generated random number sequence in IBM SPSS (Version 25). A simple randomization method was applied to ensure an equal probability of assignment. The randomization list was prepared by an independent researcher not involved in participant recruitment or data collection. Allocation concealment was maintained through sealed, opaque, and sequentially numbered envelopes prepared by a research assistant not involved in the intervention or data analysis. The envelopes were opened only after participants consented, preventing selection bias.

To minimize contamination between the EXP and CON groups, participants were drawn from different institutions and received instruction through separate online sessions conducted on different days. Communication about course materials and activities was restricted via clear instructions and monitoring of online platforms. The intention-to-treat (ITT) principle, which emphasizes that all participants are analyzed in their initially assigned groups regardless of intervention completion, adherence, or dropout, was not applicable, as all participants completed both pre- and post-tests. Although some participants missed online sessions, make-up sessions were held in person to ensure full delivery of the intervention. Consequently, all participants were included in the analysis, and no attrition or missing data necessitated the use of ITT analysis.

The EXP group took part in a structured 30-hour, 2-credit IE course designed for this study to improve their knowledge, skills, and beliefs about inclusive practices. The course content covers the foundational ideas of IE, basic identification of children with SEN by their specific traits, how they are supported in classrooms to meet their diverse needs, IE policies and laws enacted internationally and locally, inclusive teaching methods, classroom management strategies with proactive and reactive approaches, formative and informal assessment techniques with the development of rubrics, and collaboration skills for supporting students with SEN. The course content was reviewed by IE experts using a 4-point Likert-type rating scale to assess its relevance, clarity, and comprehensiveness, thereby ensuring content validity.

Conducted by the chief investigator in a hybrid format combining in-person and Zoom sessions, the course employed a flipped learning approach. Designed in accordance with the Sri Lanka Qualifications Framework, this course can be incorporated as a standalone component into ITE programs, subject to approval from the relevant authorities. Since the practicum was institutionally determined, the intervention was limited to coursework and workshops.

The instructional intervention was carried out using a flipped learning approach in three stages. Zoom-recorded video lessons and related open-ended questions were shared before class to encourage independent exploration. Participants attended synchronous sessions, prepared to discuss and apply concepts through interactive activities such as group work, role-playing, presentations, debates, and quizzes. In-class practice focused on microteaching and simulated classroom tasks that developed key skills, like the use of teaching aids, reinforcement, classroom and behaviour management, formative assessment, stimulus variation, and questioning, which were deemed crucial for meeting the diverse needs of learners. Each session included structured peer and instructor feedback to enhance instructional competence. Simulated classroom sessions were organized in small groups to provide practical experience in applying the competencies gained through microteaching, including lesson planning, preparing teaching aids, designing behaviour and classroom management strategies, and developing formative and summative assessment methods.

Fidelity checks were carried out using a checklist, online feedback from sessions, and session observations. Specifically, participants' pre-class engagement was indirectly assessed through their written reflections and their ability to respond to guiding questions during subsequent in-class discussions. Additionally, participants' reflective learning logs and Know–Want to Know–Learned (KWL) charts, completed after each lesson, were reviewed with feedback to track their engagement with pre-class learning.

Data analysis

The hypotheses were tested using IBM SPSS (version 25) and SmartPLS 4. Descriptive statistics (means, standard deviations, and internal consistencies) were calculated for all study variables. To assess baseline equivalence, pre-test scores were compared between the EXP and CON groups across all outcome variables (EII, EMB, EC, and TEIP), as well as between the online and onsite delivery groups, to verify that the groups were comparable before the intervention. Paired-samples t-tests were used to examine within-group changes from pre- to post-tests for the EXP and CON groups across all outcome variables. Post-test scores were then compared between the EXP and CON groups and between the online and onsite groups to determine whether the intervention produced significant improvements and whether delivery mode influenced outcomes.

To test the study hypotheses, two complementary analytical procedures were employed. First, a one-way MANCOVA in IBM SPSS was performed to examine whether TEIP post-test subscale scores differed between the EXP and CON groups, controlling for baseline scores. Although independent-samples *t*-tests showed no significant pre-test differences between the groups, preliminary analyses indicated that the pre-test scores of EII, EMB, and EC significantly predicted their respective post-test scores. As a result, MANCOVA was used with each post-test score as the dependent variable, group (EXP vs. CON) as the independent variable, and the relevant pre-test scores as covariates to obtain adjusted post-test means. This study favoured MANCOVA over repeated-measures ANOVA because it accounts for baseline differences, reduces error variance, and provides a more accurate and less biased estimate of the intervention effect (Field, 2018; Dimitrov & Rumrill, 2003).

Additionally, PLS-SEM was conducted in SmartPLS4 to estimate the predictive relationships between the intervention and post-test outcomes, while accounting for pre-test scores as covariates. PLS-SEM extended the analysis by validating the measurement model (outer model) of the TEIP and by assessing standardized path coefficients, effect sizes, and explained variance (R^2) in the structural model (inner model). Together, these procedures allowed the hypotheses to be tested both at the classical group-level and within a predictive modelling framework.

RESULTS

Descriptive Statistics for Outcomes Before and After the Intervention

Table 1 presents the mean scores, standard deviations, and post-intervention gains in the mean scores for the pre- and post-test data, grouped by the EXP and CON conditions, for the overall TEIP construct and its dimensions, EII, EMB, and EC. The EXP group showed significant improvements across all outcomes after the intervention, with gains of 1.92 for EII, 1.18 for EMB, 1.31 for EC, and 1.52 for the total TEIP. The CON group showed only minor changes, with gains of .53 for EII, .17 for EMB, .08 for EC, and .29 for the TEIP. Overall, the descriptive statistics indicate that gains in the EXP group were consistently larger and accompanied by moderate variability, as reflected in the standard deviations. In contrast, the CON group showed minimal changes and greater variability. This pattern indicates that the intervention had a pronounced and relatively consistent impact on participants in the EXP group, while the CON group remained considerably stable.

Table 1. *Mean Scores and Standard Deviations for Pre- and Post-Tests by Group*

Outcome	Group	N	Pre-test Mean (SD)	Post-test Mean (SD)	Mean Difference
EII	EXP	120	3.31(.307)	5.23(.296)	+1.92
	CON	120	3.32(.335)	3.85(.398)	+0.53
EMB	EXP	120	3.20(.405)	4.38(.791)	+1.18
	CON	120	3.24(.315)	3.41(.317)	+0.17
EC	EXP	120	3.07(.450)	4.38(.739)	+1.31
	CON	120	3.07(.449)	3.15(.491)	+0.08
TEIP	EXP	120	3.20(.217)	4.72(.330)	+1.52
	CON	120	3.21(.262)	3.50(.256)	+0.29

Baseline Equivalence Checks Between Groups

EXP versus CON Groups

To ensure comparability between the EXP and CON groups before the intervention, independent-samples *t*-tests were conducted on pre-test scores for EII, EMB, EC, and TEIP. Establishing baseline equivalence is essential in experimental designs, as significant pre-intervention differences can confound treatment effects and threaten internal validity (Field, 2018; Tabachnick & Fidell, 2019). By confirming that there are no significant baseline differences, post-test results can be interpreted with greater confidence (Pallant, 2020). The results indicated that the groups were equivalent at pre-test: EII, $t(238) = –0.40$, $p = .688$; EMB, $t(224.30) = –0.82$, $p = .414$; EC, $t(238) = 0.05$, $p = .962$; and total TEIP, $t(229.98) = –0.36$, $p = .721$. In all cases, the mean differences were minimal, and

the 95% confidence intervals included zero, indicating no statistically significant differences between groups at baseline. In addition to randomization, these results support that the groups were equivalent before the intervention.

Online versus Onsite Groups

To assess baseline equivalence between the online (N = 48) and onsite (N = 72) groups, an independent-samples *t*-test was conducted to compare pre-test scores for EII, EMB, EC, and TEIP separately. The results indicated no significant differences for total TEIP_PRE ($t_{(118)} = -.856$, $p = .394$), EII_PRE ($t_{(118)} = .190$, $p = .850$), or EMB_PRE ($t_{(118)} = 1.347$, $p = .181$). However, for EC, the analysis revealed a significant difference between online ($M = 2.97$, $SD = .435$) and onsite ($M = 3.14$, $SD = .450$) modes of delivery ($t_{(118)} = -2.127$, $p = .036$), indicating that the onsite group reported slightly higher levels of EC prior to the intervention. However, the magnitude of this difference was small, and subsequent PLS-SEM analyses accounted for baseline EC differences, reducing the likelihood that this imbalance unduly influenced the intervention effects. Nevertheless, this factor should be considered when interpreting the findings.

Post-Intervention Checks Between Groups

Online versus Onsite Groups

To ensure that any observed effects could be attributed to the intervention rather than the delivery mode, an independent-samples *t*-test was conducted to compare post-test TEIP scores across its three dimensions and the overall TEIP between the online (N = 48) and onsite (N = 72) groups. With a statistically significant Levene's test, which indicated that equal variances between the online and onsite delivery groups for EII post-test scores were not assumed, $F_{(1,238)} = 13.338$, $p < .001$, the *t*-test results showed a non-significant difference between the groups, $t_{(73.43)} = -0.645$, $p = .521$, indicating that delivery mode did not influence EII outcomes.

Assumption of equal variances was satisfied for EMB post-test scores, $F_{(1,238)} = 1.149$, $p = .286$, as shown by the independent-samples t-test, which indicated no significant difference between the online and onsite delivery groups, $t_{(118)} = 1.738$, $p = .085$, suggesting that the mode of delivery did not significantly influence EMB outcomes. Similarly, the F statistic for EC post-test scores confirmed that the assumption of equal variances was met, $F_{(1, 238)} = 1.071$, $p = .303$.

The independent-samples t-test results showed no significant difference between the online and onsite delivery groups, $t_{(118)} = -0.608$, $p = .544$, indicating that delivery mode did not significantly affect EC outcomes. Levene's test supported the assumption of equal variances for TEIP post-test scores, $F_{(1, 238)} = 0.239$, $p = .626$, with an independent-samples t-test, which showed no significant difference between the online and onsite delivery groups, $t_{(118)} = 0.388$, $p = .699$, suggesting that delivery mode did not influence TEIP outcomes.

Overall, the delivery mode did not influence post-test outcomes, suggesting that participants' performance was consistent regardless of whether the IE course was delivered online or onsite. Therefore, for SEM, it is reasonable and statistically justified to collapse the delivery modes and analyze the intervention group as a whole, simplifying the model and focusing on the effect of the IE course itself.

EXP versus CON Groups

Post-intervention gains in the outcome variables within the EXP group were analyzed separately using paired-samples t-tests. The results revealed significant improvements: EII increased by 1.53 points, $t_{(119)} = 42.556$, $p < .001$; EMB increased by 1.06 points, $t_{(119)} = 21.273$, $p < .001$; EC increased by 1.17 points, $t_{(119)} = 35.896$, $p < .001$; and overall TEIP advanced by 1.27 points, $t_{(119)} = 53.591$, $p < .001$. These findings demonstrate that the intervention effectively improved participants' skills and perceptions.

In comparison, the CON group showed negligible and statistically significant gains in the mean scores with EII increased by 0.37 points, $t_{(119)} = 10.963$, $p < .001$; EMB by 0.17 points, $t_{(119)} = 6.121$, $p < .001$; EC by 0.09 points, $t_{(119)} = 3.465$, $p = .001$; and TEIP by 0.22 points, $t_{(119)} = 11.678$, $p < .001$. These results indicate that while the CON group experienced minor improvements, the intervention produced substantially larger gains in the EXP group.

Post-test comparisons between the EXP and CON groups were conducted using independent-samples *t*-tests. Levene's tests indicated that the assumption of equal variances was violated for EII and EMB ($p < .001$) but met for TEIP and EC ($p > .05$). Therefore, Welch's *t*-test was used where necessary, indicating significant differences between the EXP and CON groups across all outcome variables: TEIP, $t_{(238)} = 30.776$, $p < .001$, mean difference = 1.04; EII, $t_{(213.30)} = 24.949$, $p < .001$, mean difference = 1.14; EMB, $t_{(162.38)} = 13.090$, $p < .001$, mean difference

= 0.86; and EC, $t_{(238)} = 15.538$, $p < .001$, mean difference = 1.08. These results indicate that the intervention group outperformed the control group on all post-test measures.

Assumption Checks for MANCOVA

The normality of the post-test scores for all variables was evaluated using the Shapiro–Wilk and Kolmogorov–Smirnov tests, both of which were significant ($p < .05$); however, with large sample sizes (N = 120 per group), such significance is expected because these tests are highly sensitive and can detect trivial deviations from normality (Ghasemi & Zahediasl, 2012). An objective review of histograms and Q–Q plots showed that both groups had approximately normal distributions, with no extreme skewness or kurtosis. Since MANCOVA is robust to moderate breaches of normality assumptions, especially with large, balanced groups, the results support the assumption of approximate normality, allowing the analysis to proceed without data transformation or non-parametric alternatives (Tabachnick & Fidell, 2019).

The homogeneity of regression slopes was initially tested using a GLM by examining interactions between the independent variable (EXP_CONT) and each covariate (EII-pre, EMB-pre, EC-pre) for each dependent variable. The results showed that this assumption was violated for EII-post and EMB-post, with significant interactions observed (EII-post: EXP_CONT × EII-pre, $F_{(1,236)} = 17.24$, $p < .001$; EXP_CONT × EMB-pre, $F_{(1,236)} = 8.201$, $p = .005$; EMB-post: EXP_CONT × EMB-pre, $F_{(1,236)} = 8.201$, $p = .005$), while it was met for EC-post (EXP_CONT × EC-pre, $F_{(1,236)} = 3.525$, $p = .062$). Because regression slope homogeneity was not consistent across all outcomes, a standard MANCOVA could not be used. Although ANCOVA could be applied to EC-post, the analysis was transitioned to PLS-SEM to ensure methodological consistency across all variables and to account for baseline differences observed only for EC in the delivery method.

Main Effect of the IE Course Intervention on the EII, EMB, and EC

Hypotheses that the intervention would positively predict EII, EMB, and EC were tested using PLS-SEM in SmartPLS 4 with a non-parametric bootstrapping procedure (10,000 resamples) and two-tailed t-tests, generating robust standard errors and confidence intervals for the path coefficients. The PLS-SEM analysis allows for the simultaneous modelling of multiple dependent variables, covariates, and their interactions without assuming slope homogeneity. This offers a robust framework for examining the direct effects of the IE course on EII, EMB, EC, and the overall TEIP, using two structural models—one with covariates and one without—where the intervention (EXP = 0, CON = 1) is the independent variable and post-test scores are outcomes, controlling for pre-test scores for each construct.

An initial PLS-SEM analysis was performed by excluding pre-test scores as covariates to assess the direct, unadjusted effect of the intervention on post-test outcomes, establishing a baseline estimate before considering potential confounders. The path model and coefficients are illustrated in Figure 1. Results showed that the intervention significantly improved all three dimensions of self-efficacy. With the EXP group coded as '0' and the CON group as '1', negative coefficients indicate better performance by the EXP group: EII, $B = –1.713$, $t = 64.490$, $p < .001$, 95% CI [–1.757, –1.649]; EMB, $B = –1.296$, $t = 18.295$, $p < .001$, 95% CI [–1.425, –1.147]; EC, $B = –1.423$, $t = 19.818$, $p < .001$, 95% CI [–1.550, –1.264]. These findings confirm that the IE course statistically and significantly predicted PST's TSE across its three dimensions: EII, EMB and EC.

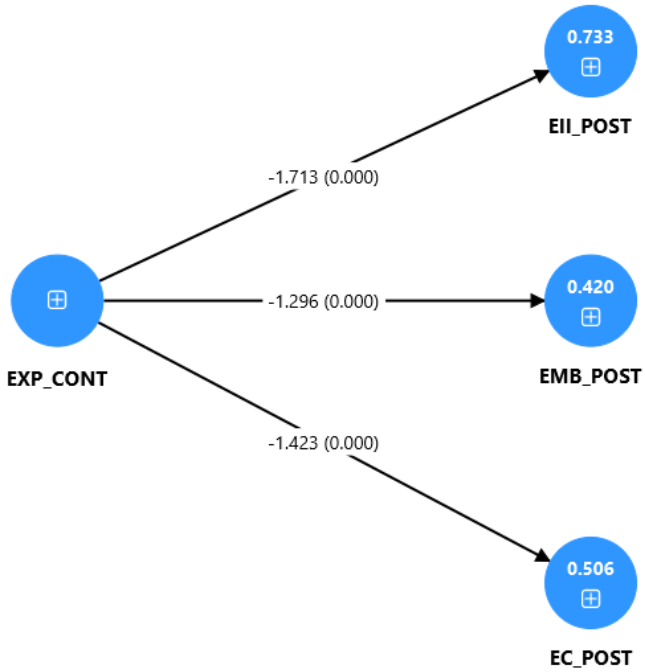


Figure 1. *The Path Model Excluding the Pre-Test Scores as Covariates*

Including pre-test scores as covariates in the PLS-SEM model showed (Figure 2) that the intervention had a significant positive effect on all TEIP dimensions, estimating the net impact adjusted for initial differences. With the EXP group coded as '0' and the CON group as '1', negative unstandardized coefficients indicate higher post-test scores for the EXP group: EII ($B = –1.676$, $t = 41.138$, $p < .001$, 95% CI [–1.779, –1.607]), EMB ($B = –1.333$, $t = 20.204$, $p < .001$, 95% CI [–1.468, –1.206]), and EC (B = –1.416, $t = 24.286$, $p < .001$, 95% CI [–1.530, –1.301). Presenting both adjusted and unadjusted results enhances transparency and interpretation of the intervention's effect.

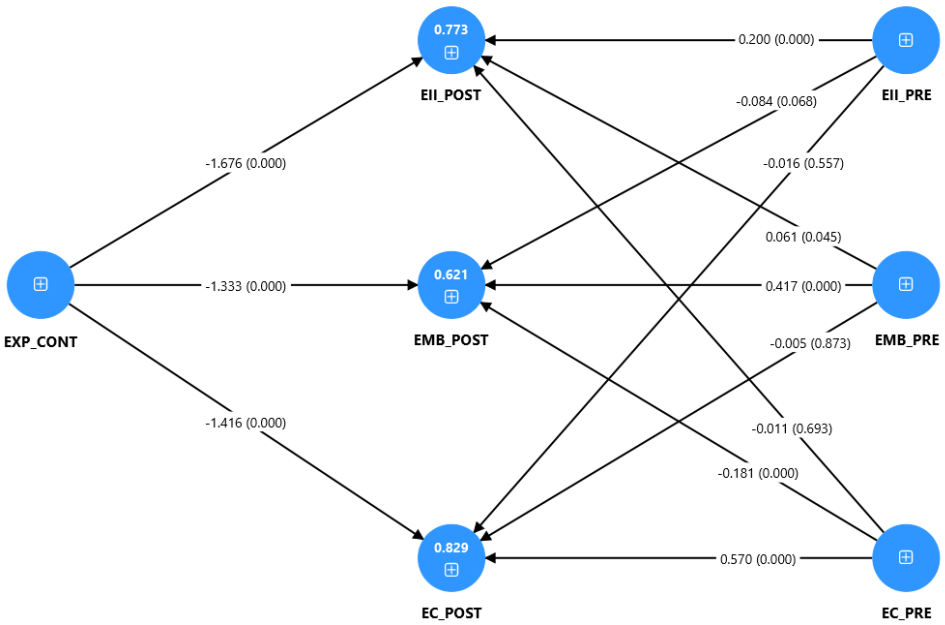


Figure 2. *The Path Model with the Pre-Test Scores Functioned as Covariates*

Furthermore, the unstandardized path coefficients indicated that pre-test scores significantly predicted their respective post-test scores, demonstrating baseline stability of each construct (Table 2). Specifically, EII-pre positively and significantly predicted EII-post ($B = 0.200$, $t = 4.673$, $p < .001$, 95% CI [0.099, 0.263]), EMB-pre positively and significantly predicted EMB-post ($B = 0.417$, $t = 8.606$, $p < .001$, 95% CI [0.314, 0.504]), and EC-

pre positively and significantly predicted EC-post ($B = 0.570$, $t = 12.295$, $p < .001$, 95% CI [0.476, 0.657]). These findings demonstrate that higher baseline levels of each construct were linked to higher post-test scores for the same construct, after controlling for the intervention.

Table 2. *Path Coefficients, Significance, and 95% BCa Confidence Intervals for the Effects of the IE Course and Pre-Test Scores on Post-Test Self-Efficacy Outcomes*

Path	*B*(O)	*M*	*SD*	*t*	*p*	95%CI(BCa)
Intervention Effect						
GROUP → EII_POST	–1.676	-1.673	0.041	41.14	.000	-1.779, -1.607
GROUP → EMB_POST	–1.333	-1.333	0.066	20.20	.000	-1.468, -1.206
GROUP → EC_POST	–1.416	-1.418	0.058	24.29	.000	-1.530, -1.301
Covariate Effects of Their Own Post-Test Scores						
EII_PRE → EII_POST	0.200	0.196	0.043	4.67	.000	0.099, 0.263
EMB_PRE → EMB_POST	0.417	0.416	0.048	8.61	.000	0.314, 0.504
EC_PRE → EC_POST	0.570	0.570	0.046	12.30	.000	0.476, 0.657
Cross-Construct Pre-Test Effects						
EC_PRE → EII_POST	–0.011	-0.012	0.029	0.39	.693	-0.070, 0.045
EC_PRE → EMB_POST	–0.181	-0.180	0.033	5.57	.000	-0.247,-0.120
EII_PRE → EC_POST	–0.016	-0.015	0.027	0.59	.557	-0.070, 0.037
EII_PRE → EMB_POST	–0.084	-0.080	0.046	1.83	.068	-0.169, 0.010
EMB_PRE → EC_POST	–0.005	-0.004	0.032	0.16	.873	-0.068, 0.056
EMB_PRE → EII_POST	0.061	0.059	0.030	2.01	.045	0.003, 0.122

Some cross-construct effects of pre-test scores were also observed: EC-pre negatively predicted EMB-post ($B = –0.181$, $t = 5.57$, $p < .001$, 95% CI [–0.247, –0.120]), and EMB-pre positively predicted EII-post ($B = 0.061$, $t = 2.01$, $p = .045$, 95% CI [0.003, 0.122]), suggesting that pre-existing levels of one construct had a small but significant influence on changes in another construct. Other cross-construct effects of pre-test scores (e.g., EC-pre → EII-post, EII-pre → EC-post) were not significant, indicating limited spill-over effects. Overall, these findings confirm that the intervention, combined with the IE course, significantly improved PSTs' TSE in inclusive instruction, behaviour management, and collaboration, even after controlling for baseline scores.

When comparing the effects of the IE course on post-test outcomes with and without controlling for pre-test scores, it is clear that the intervention significantly improved all three dimensions of TSE in both analyses. Without covariates, the effects were slightly stronger. After including pre-test scores as covariates, the effects remained significant but were slightly smaller. This comparison indicates that part of the intervention's effect overlaps with baseline differences in TSE. However, the IE course still shows a significant impact, even after accounting for baseline influences. The slight decrease in effect sizes after covariate adjustment suggests that pre-test scores account for only a small part of the post-test variance. Nonetheless, the intervention itself remains the primary factor affecting post-test TSE.

DISCUSSION

As part of a larger project examining the mediating role of PSTs' beliefs about IE in the relationship between completing a structured IE course delivered through a flipped learning approach and the TEIP dimensions (EII, EMB, and EC), this article focuses solely on the intervention's direct impact. The hypotheses that the intervention would demonstrate predictive effects on the mean post-test gain scores for EII, EMB, EC, and TEIP, adjusted for pre-test differences, were supported, specifically, with the EII dimension demonstrating the highest unstandardized path coefficients and being most strongly influenced by the intervention.

These findings align with Bandura's (1977, 1997) sources of efficacy principle, where the mastery experiences, gained through multiple microteaching experiences, and practicing inclusive lesson plans, vicarious experiences, gained through observing their peers successfully performing teaching in simulated classrooms and obtaining 360-degree constructive feedback from their co-participants and the instructor, together contributed to the development of TSE. Previous empirical findings further support the findings. For example, Lancaster and Bain (2007) found that PSTs' TSE improved after an IE course regardless of the type of experience. However, Can (2015) emphasized that both school-based experiences and faculty-led courses are key components in enhancing TSE.

Notably, in the present study, significant gains were observed in EII, EMB, and EC, despite the intervention being delivered entirely through coursework, including simulated teaching and microteaching, without a practicum component. This diverges from Bandura's emphasis on mastery experiences as the primary source of efficacy, yet

aligns with Lancaster and Bain, suggesting that well-designed simulated coursework can effectively enhance PSTs' confidence in inclusive instruction.

Nevertheless, there are contradicting outcomes regarding the impact of the IE course on TSE. Tait and Purdie (2000) found that a 12-month teacher training course did not change PSTs' views on disabilities and inclusion. Stella et al. (2007) reported that even after completing a brief instructional module on inclusive philosophy and practices, participants' attitudes toward inclusion changed very little. Importantly, Nagata (2005) argues that a single subject on inclusion cannot adequately prepare new teachers to handle the many tasks involved in inclusive practice or meet the demands of an inclusive classroom.

These different patterns of outcomes suggest that the effectiveness of IE courses heavily depends on their design and the practical experiences they provide. For many authors, brief modules or single-subject courses may be insufficient to bring about meaningful changes in PSTs' TSE or attitudes towards inclusion (Nagata, 2005; Stella et al., 2007; Tait & Purdie, 2000). Conversely, a well-structured coursework with simulated and microteaching, along with guided reflection, as implemented in the present study, can substantially boost confidence in PSTs' abilities in inclusive practices, even without a field-based practicum (Mergler & Tangen, 2010). This emphasizes that meaningful, practice-oriented experiences, rather than mere content exposure, are essential for developing practical, inclusive teaching skills.

The study also found that the intervention with the IE course was a significant predictor of each TEIP factor. The most notable improvements were in the EII, while the smallest were in the EMB, indicating that the course effectively increased participants' confidence in using inclusive instruction. This varied impact of the intervention suggests that PSTs showed greater confidence in their abilities in inclusive instructional planning and design, including the implementation of Universal Design for Learning, differentiated instruction, and backward design. Additionally, their confidence in applying inclusive practices, such as providing accommodations and making curriculum modifications, also appeared to be enhanced after the intervention. Notably, these improvements were evident following the course-based learning experience delivered through simulated environments, despite participants not having direct teaching experience in actual inclusive classrooms.

This finding aligns with previous research, suggesting that structured IE training tends to have a greater influence on instructional efficacy than on other domains (Forlin et al., 2014). Gains in EMB and EC were more modest (+1.18 and +1.31, respectively), though still exceeding those of the CON group (+0.17 and +0.08). These comparatively lower gains reflect the greater complexity involved in developing behavioural management and collaborative competencies, which often require extended practice and systemic support. Overall, the pattern suggests that short-term interventions are more effective in strengthening instructional efficacy than in fostering behavioural or collaborative capacities, aligning with international evidence on the domain-specific development of TSE (Loreman et al., 2013).

The findings further extended Bandura's (1997) self-efficacy framework by illustrating how course delivery methods, such as a flipped learning approach, combined with simulation, can operationalize the sources of efficacy information. Simulated environments created quasi-mastery experiences by enabling participants to practice inclusive teaching in realistic yet low-risk contexts. Peer modelling provided vicarious experiences, while feedback from peers, the instructor and reflection functioned as verbal persuasion, reinforcing participants' confidence. Together, these elements offer a contemporary model for developing TSE to support inclusive practices through experiential, learner-centered pedagogy.

Although the control group showed small but statistically significant increases from pre- to post-test, this pattern is expected in educational studies. Participants, in general, gain confidence as they progress through their regular coursework, become more familiar with teaching concepts, or become used to the measurement tool. Such maturational or testing effects can produce moderate improvements even in the absence of a targeted intervention. Importantly, these gains were comparatively negligible compared to those observed in the intervention group, indicating that the substantial increases in TSE were driven by the IE course intervention rather than general program exposure.

SmartPLS reports unstandardized path coefficients in the structural model, which explains the large numerical values observed in this study. These coefficients depend on the original metric and the variance of each latent variable. Therefore, constructs with minor variances will naturally produce larger unstandardized coefficients. SmartPLS does not automatically generate standardized coefficients *(β)* in the structural model output; instead, standardization is typically reflected only in measurement model loadings. Therefore, the large unstandardized

values should not be interpreted as powerful effects. Standardized coefficients—when computed—would provide a more comparable indication of effect size across constructs.

The current findings have several implications. First, the course design and delivery, including the use of flipped learning and a learner-centered approach, provide empirical evidence of a practical method for enhancing PSTs' confidence in implementing inclusive practices. In curricular reforms, such findings suggest that structured IE courses with such delivery approaches can be feasibly integrated into ITE programs without overloading schedules, providing PSTs with both theoretical knowledge and practical skills to support diverse learners effectively. Second, from the policy perspective, the study supports the intensive implementation of flipped learning-oriented IE courses, highlighting the need for faculty training and resources to implement these pedagogies, thereby contributing to more inclusive teaching practices nationwide. Finally, the results encourage future research on mediating factors, such as PST beliefs and long-term retention of TSE, and comparisons with traditional lecture-based IE courses to optimize teacher preparation for inclusive classrooms further.

LIMITATIONS AND RECOMMENDATIONS

The interpretation of the findings of this study should be approached with caution in light of the following limitations. Its findings may have limited generalizability to the broader population of PSTs in Sri Lanka, as it employed only Tamil-speaking PSTs. The short duration of the flipped learning course may not have allowed enough time for lasting changes in TSE to develop. Additionally, reliance on self-reported quantitative measures may have introduced social desirability and response biases, and novelty effects linked to the instructional approach may have temporarily elevated confidence. Most importantly, participants' inconsistent attendance and engagement during online sessions may have limited the extent to which they benefited fully from the intervention.

Despite random assignment, other extraneous variables, such as prior exposure to IE concepts, children with SEN or institutional learning culture, could have influenced participants' TSE development. Diffusion of treatment across institutions may also have reduced group distinctions. These factors call for a cautious interpretation of the findings and highlight the need for more robust research designs in future studies. In addition, the lack of delayed post-tests restricts conclusions about the sustainability of the outcomes, as the post-test was conducted immediately after the intervention. Consequently, it remains unclear whether the observed increases in TSE were maintained over time.

Although the IE course demonstrated a significant positive impact on TSE for inclusive practices within a longitudinal design, several methodological refinements are recommended for future research. First, future studies could employ multilevel modelling to account for clustering within ITE institutions. This effort would enable more precise estimates of the intervention effects at both the individual and institutional levels. Second, adopting longitudinal designs with delayed post-tests would help determine whether gains in TSE are sustained during practicum or early teaching. Third, it is recommended to adopt qualitative research approaches, including phenomenological interviews, focus group discussions, classroom observations, or behavioural assessments, as complementary measures of self-report, to rule out the effects of social desirability and response bias on the outcomes, thereby triangulating and strengthening the validity of the findings.

CONCLUSIONS

The study found that the IE course intervention significantly enhanced PSTs' TSE in inclusive instructional practices, behaviour management, and collaboration, with the most potent effect on instructional practices. The structured, theory-based, modular IE course, delivered through a flipped learning approach without a practicum, directly improved TSE. Using an experimental pre- and post-test control-group design ensured a rigorous comparison, confirming the intervention as a significant predictor of TSE gains across all teacher efficacy dimensions.

Author Contributions

The principal author, Y.S., is responsible for all aspects of the study, including the development of the IE course, conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, drafting and revising the manuscript, visualization, and project administration, with the co-author, K. A. C., providing supervision.

Declarations

Competing interests

The authors declare that they have no competing interests.

REFERENCES

Alazemi, A., & Larkins, C. (2025). Teachers' perceptions of preparedness for teaching students with mild to moderate disabilities in inclusive education in Kuwait. *International Journal of Special Education, 40*(1), 1–16. https://doi.org/10.52291/ijse.2025.40.1

Avramidis, E., Bayliss, P., & Burden, R. (2000). Student teachers' attitudes towards the inclusion of children with special educational needs in the ordinary school. *Teaching and Teacher Education, 16*(3), 277–293. https://doi.org/10.1016/S0742-051X(99)00062-1

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215.https://doi.org/10.1037/0033-295X.84.2.191

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York: W.H. Freeman and Company.

Brouwers, A., & Tomic, W. (2000). A longitudinal study of teacher burnout and perceived self-efficacy in classroom management. *Teaching and Teacher Education*, *16*, 239–253. https://doi.org/10.1016/S0742-051X(99)00057-8

Can, H. (2015). *Sources of teaching efficacy beliefs in pre-service science teachers*. İlköğretim Online, 14(3), 1044-1059. https://doi.org/10.17051/io.2015.44466

Carroll, A., Forlin, C., & Jobling, A. (2003). *The impact of teacher training in special education on the attitudes of Australian preservice general educators towards people with disabilities*. *Teacher Education Quarterly, 30*(3), 65–79.

Casarez, L. L. (2013). *Pre-service teachers' attitudes toward students with disabilities: A study of candidates enrolled in teacher preparation programs.* https://bit.ly/3PzN2wv

Dimitrov, D. M., & Rumrill, P. D. (2003). Pretest–posttest designs and measurement of change. *Work, 20*(2), 159–165. https://doi.org/10.3233/WOR-2003-00285

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). Sage.

Forlin, C. (2010). Teacher education reform for enhancing teachers' preparedness for inclusion. *International Journal of Inclusive Education, 14*(7), 649–653. https://doi.org/10.1080/13603111003778353

Forlin, C., & Chambers, D. (2011). Teacher preparation for inclusive education: Increasing knowledge but raising concerns. *Asia-Pacific Journal of Teacher Education, 39*(1), 17–32. https://doi.org/10.1080/1359866X.2010.540850

Forlin, C., Sharma, U., & Loreman, T. (2014). Predictors of improved teaching efficacy following basic training for inclusion in Hong Kong. *International Journal of Inclusive Education, 18*(7), 718–730. https://doi.org/10.1080/13603116.2013.819941

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: A guide for non-statisticians. *International Journal of Endocrinology and Metabolism, 10*(2), 486–489. https://doi.org/10.5812/ijem.3505

Hair, J.F., Black, W.C., Babin, B.J., & Anderson, R.E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.

Hemmings, B., & Weaven, E. (2005). *The reality of undertaking an inclusive education internship*. In F. Bryer (Ed.), *Making meaning: Creating connections that value diversity* (pp. 86–93). Australian Association of Special Education (AASE).

Hettiarachchi, S., & Das, A. (2014). Perceptions of 'inclusion' and perceived preparedness among school teachers in Sri Lanka. *Teaching and Teacher Education, 43*, 143-153. https://doi.org/10.1016/j.tate.2014.07.003

Kurniawati, F., de Boer, A. A., Minnaert, A. E. M. G., & Mangunsong, F. (2014). Characteristics of primary teacher training programmes on inclusion: A literature focus. *Educational Research, 56*(3), 310–326. https://doi.org/10.1080/00131881.2014.934555

Lambe, J., & Bones, R. (2007). The effect of school-based practice on student teachers' attitudes towards inclusive education in Northern Ireland. *Journal of Education for Teaching, 33*(1), 99–113. https://doi.org/10.1080/02607470601098369

Lancaster, J., & Bain, A. (2007). The design of inclusive education courses and the self-efficacy of pre-service teacher education students. *International Journal of Disability, Development and Education, 54*(2), 245–256. https://doi.org/10.1080/10349120701330610

Lancaster, J., & Bain, A. (2010). The design of pre-service inclusive education courses and their effects on self-efficacy: A comparative study. *Asia-Pacific Journal of Teacher Education, 38*(2), 117–128. https://doi.org/10.1080/13598661003678950

Loreman, T., Sharma, U., & Forlin, C. (2013). Do pre-service teachers feel ready to teach in inclusive classrooms? A four country study of teaching self-efficacy. *Australian Journal of Teacher Education, 38*(1). https://doi.org/10.14221/ajte.2013v38n1.10

Lucena-Rodríguez, I., Suárez-Martín, E., & Rodríguez-Martín, A. (2025). Deficiencies in the academic preparation of future teachers for inclusive education: A systematic review. *International Journal of Instruction, 18*(1), 199–216. https://doi.org/10.29333/iji.2025.1834a

Mergler, A., & Tangen, D. (2010). *Using microteaching to enhance teacher efficacy in pre-service teachers. Teaching Education, 21*(2), 199–210. https://doi.org/10.1080/10476210902998466

Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gøtzsche, P. C., Devereaux, P. J., Elbourne, D., & Egger, M.; Altman, D. G. (2010). CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *Journal of Clinical Epidemiology, 63*(8), e1–e37. https://doi.org/10.1016/j.jclinepi.2010.03.004

Nagata, N. (2005). *Characteristics of teacher preparation programs and the issue perceptions of teacher educators in deaf education* [Unpublished doctoral dissertation]. The Ohio State University, Columbus.

Nketsia, W., & Saloviita, T. (2013). Pre-service teachers' views on inclusive education in Ghana. *Journal of Education for Teaching: International Research and Pedagogy, 39*(4), 429–441. https://doi.org/10.1080/02607476.2013.797291

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

O'Neill, S. C. (2015). Preparing preservice teachers for inclusive classrooms: Does completing coursework on managing challenging behaviours increase their classroom management sense of efficacy? *Australasian Journal of Special Education. Advance online publication.* https://doi.org/10.1017/jse.2015.10

Pallant, J. (2020). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS* (7th ed.). McGraw-Hill Education. https://doi.org/10.4324/9781003117452

Pendergast, D., Garvis, S., & Keogh, J. (2011). Pre-service student-teacher self-efficacy beliefs: An insight into the making of teachers. *Australian Journal of Teacher Education, 36*(12), 46–58. https://doi.org/10.14221/ajte.2011v36n12.6

Peries, A. N., Indrarathne, B., Jayamanne, B. D. W., Wickramasekara, T. D., Alwis, K. A. C., & Jayatilleke, A. U. (2021). Primary school teachers' readiness in identifying children with dyslexia: A national survey in Sri Lanka. *Dyslexia, 27*(4), 486–509. https://doi.org/10.1002/dys.1696

Sakthivel, Y. (in press). Cross-cultural adaptation and validation: Teacher Efficacy for Inclusive Practices scale for Tamil-speaking pre-service teachers. *International Journal of Assessment Tools in Education.*

Schulz, K. F., & Grimes, D. A. (2002). Allocation concealment in randomised trials: Defending against deciphering. The Lancet, 359(9306), 614–618. https://doi.org/10.1016/S0140-6736(02)07750-4

Sharma, U. (2012). Changing pre-service teachers' beliefs to teach in inclusive classrooms in Victoria, Australia. *Australian Journal of Teacher Education, 37*(10), 53–66. https://doi.org/10.14221/ajte.2012v37n10.6

Sharma, U., & Nuttall, A. M. (2016). The impact of training on pre-service teacher attitudes, concerns, and efficacy towards inclusion. *Asia-Pacific Journal of Teacher Education, 44*(2), 142–155. https://doi.org/10.1080/1359866X.2015.1081672

Sharma, U., & Sokal, L. (2015). The impact of a teacher education course on pre-service teachers' beliefs about inclusion: An international comparison. *Journal of Research in Special Educational Needs, 15*(4), 276–284. https://doi.org/10.1111/1471-3802.12043

Sharma, U., Forlin, C., & Loreman, T. (2008). Impact of training on pre-service teachers' attitudes and concerns about inclusive education and sentiments about persons with disabilities. *Disability & Society, 23*(7), 773–785. https://doi.org/10.1080/09687590802469271

Sharma, U., Forlin, C., & Simi, J. (2015). Preparedness of pre-service teachers for inclusive education in the Solomon Islands. *Australian Journal of Teacher Education, 40*(5), 64–79. https://doi.org/10.14221/ajte.2015v40n5.6

Sharma, U., Loreman, T., & Forlin, C. (2012). Measuring teacher efficacy to implement inclusive practices. *Journal of Research in Special Educational Needs, 12*(1), 12–21. https://doi.org/10.1111/j.1471-3802.2011.01200.x

Stella, C. S. C., Forlin, C., & Lan, A. M. (2007). The influence of an inclusive education course on attitude change of pre-service secondary teachers in Hong Kong. *Asia-Pacific Journal of Teacher Education, 35*(2), 161–179. https://doi.org/10.1080/13598660701268585

Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson

Tait, K., & Purdie, N. (2000). Attitudes towards disability: Teacher education for inclusive environments in an Australian university. International Journal of Disability, Development and Education, 47(1), 25–38. DOI:10.1080/103491200116110

Torgerson, D. J., & Torgerson, C. J. (2008). *Designing randomised trials in health, education and the social sciences*. Palgrave Macmillan. https://doi.org/10.1057/9780230583993

Tschannen-Moran, M., & Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, 17*, 783–805. https://doi.org/10.1016/S0742-051X(01)00036-1

UNESCO. (2008). *Inclusive education: The way of the future: Reference document*. International Conference on Education, 48th session, Geneva, 25–28 November 2008. UNESCO International Bureau of Education. https://unesdoc.unesco.org/ark:/48223/pf0000183277

United Nations Children's Fund (UNICEF). (2013). *The state of the world's children 2013: Children with disabilities*. UNICEF. https://doi.org/10.18356/6a978726-en

Woodcock, S., Hemmings, B., & Kay, R. (2012). Does study of an inclusive education subject influence pre-service teachers' concerns and self-efficacy about inclusion? *Australian Journal of Teacher Education, 37*(6), Article 1. https://doi.org/10.14221/ajte.2012v37n6.5

Woolfolk Hoy, A. (2000, April). *Changes in teacher efficacy during the early years of teaching.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Woolfolk Hoy, A., & Spero, R. B. (2005). Changes in teacher efficacy during the early years of teaching: A comparison of four measures. *Teaching and Teacher Education, 21*(4), 343–356. https://doi.org/10.1016/j.tate.2005.01.007

Yogaranee, S. (2024). Perspectives of pre-service and in-service general education teachers of the Faculty of Education on inclusive education: A phenomenological study. *Research Journal of Educational Sciences, 12*(2), 1–14. http://bit.ly/456x0xa

Yogaranee, S. (2025). How do pre-service teachers perceive their teacher education courses? The impact on inclusive classroom management self-efficacy. *Journal of Education and Training Studies, 13*(1), 1–24. https://doi.org/10.11114/jets.v13i1.7271

Zee, M., & Koomen, H. M. Y. (2016). Teacher self-efficacy and its effects on classroom processes, student academic adjustment, and teacher well-being: A synthesis of 40 years of research. *Review of Educational Research, 86*(4), 981–1015. https://doi.org/10.3102/0034654315626801

Zundans-Fraser, L., & Lancaster, J. (2012). Enhancing the inclusive self-efficacy of pre-service teachers through embedded course design. *Education Research International, 2012*, Article 581352. https://doi.org/10.1155/2012/581352

Enhancing Educational Quality with Explainable AI: Interpretable Prediction of Student Success

Ahmet ADALIER[1], Kian JAZAYERI[1], Moein JAZAYERI[2], Damla KARAGOZLU[1]
[1] *Department of Management Information Systems, Cyprus International University, Nicosia, North Cyprus.*
[2] *Department of Electrical and Electronics Engineering, Cyprus International University, Nicosia, North Cyprus.*
aadalier@ciu.edu.tr, kjazayeri@ciu.edu.tr, mjazayeri@ciu.edu.tr, dkaragozlu@ciu.edu.tr

Abstract
Ensuring academic success and educational equity has become a critical priority for higher education institutions worldwide. This study investigates the determinants of academic success through interpretable machine learning and explainable artificial intelligence techniques. A dataset comprising 80,000 students was utilized to develop multiple regression models for GPA prediction, followed by SHAP dependence plots and Individual Conditional Expectation (ICE) plots to examine both global and instance-level contributions of behavioral, psychological, and environmental factors. Local interpretability was further explored through SHAP waterfall plots for selected students. Results indicate that study hours, motivation level, stress, sleep duration, study environment, and access to tutoring exert particularly strong influences on academic outcomes. The findings underscore that educational quality hinges on holistic student well-being beyond curriculum design alone. By delivering transparent, evidence-based explanations at both aggregate and individual levels, this research advances United Nations Sustainable Development Goal 4 (Quality Education) and offers higher education institutions actionable, data-driven guidance for equitable, student-centered enhancement strategies.
Keywords - Explainable Artificial Intelligence, Student Success, Learning Analytics, Higher Education Quality, Quality Education

1. Introduction

In recent years, the pursuit of academic excellence in higher education has been increasingly supported by data-driven approaches (Rodríguez-Ortiz et al., 2025). The availability of rich student data, combined with advances in machine learning (ML) and explainable artificial intelligence (XAI), has enabled educators and researchers to better understand the multifaceted factors influencing student performance (López-Meneses et al., 2025). However, the application of these technologies has often been limited to predictive accuracy, with less attention given to the interpretability and educational value of the models. To ensure meaningful and equitable educational improvement, models that not only predict but also explain learning outcomes (at both global and local levels) are essential.

The current study was designed to address this need by developing an interpretable ML-based framework capable of identifying the most influential behavioral, psychological, and environmental factors affecting students' academic performance. Using a large-scale dataset of 80,000 students that includes demographic, lifestyle, and learning habit indicators, the study employs SHapley Additive exPlanations (SHAP) dependence plots, Individual Conditional Expectation (ICE) plots, and local SHAP waterfall explanations to uncover how each feature contributes to variations in GPA and to illustrate individual-level prediction mechanisms. By prioritizing explainability over mere prediction, this approach aligns with the goals of evidence-based decision-making and transparency in educational quality enhancement.

This research contributes directly to United Nations Sustainable Development Goal 4 (Quality Education) by providing insights that promote data-informed educational interventions and institutional improvement. Through explainable modeling, the findings enable educators and policy-makers to design targeted strategies that nurture academic success, psychological well-being, and learning equity. Moreover, the study exemplifies how AI can be responsibly and ethically integrated into higher education systems to support continuous quality assurance and student-centered development.

2. Literature Review

Learning analytics (LA) and machine learning (ML) have emerged as central components of the data-driven transformation of higher education. Over the past decade, large-scale digitalization of learning environments and institutional information systems has generated vast quantities of student data, opening opportunities for predictive modeling, personalized learning, and quality enhancement (Sciarrone, 2018). The convergence of LA, educational data mining (EDM), and ML has enabled researchers to go beyond basic descriptive statistics towards modeling

complex behavioral, psychological, and contextual factors that shape learning outcomes. Within this paradigm, LA is viewed not merely as an evaluative tool but as an evidence-based mechanism for enhancing teaching, learning, and institutional decision-making.

Early work in LA primarily concentrated on integrating computational techniques from statistics, human–computer interaction, and data science to monitor and interpret learning processes (Gasevic et al., 2014). These studies emphasised two persistent challenges: the lack of analytics methods native to the educational domain and the limited reflection of social or pedagogical dimensions within computational models. The subsequent development of LA research addressed these issues by embedding ML algorithms into educational contexts to enhance prediction and intervention capabilities. Such efforts have yielded robust models for student classification, dropout prediction, and performance forecasting, which are core applications that remain central to modern educational analytics (Vinoth Kumar et al., 2025).

Parallel to methodological advances, the philosophical and epistemological foundations of learning analytics have also been re-examined. Doroudi (2024) argued that the field must adopt a constructivist epistemology that recognizes learning as a socially situated and interpretive process rather than a purely statistical phenomenon. Doroudi positioned different LA approaches along the bias–variance trade-off, proposing that the next generation of analytics should balance predictive precision with conceptual interpretability. This perspective aligns with the current movement towards transparent, explainable, and pedagogically grounded AI models.

Systematic reviews provide strong empirical backing for this methodological advancement. Rodríguez-Ortiz et al. (2025) conducted a comprehensive synthesis of 101 studies published between 2018 and 2025, demonstrating that ML predominantly drives current LA applications, especially in engagement prediction, dropout modelling, and academic performance forecasting. However, the review also indicated that generative AI (GenAI) techniques, such as transformer-based architectures including GPT-4 and BERT, are emerging in adaptive feedback and sentiment analysis but remain largely experimental due with transparency and implementation issues. Similarly, Ersozlu et al. (2024) examined 77 studies and found that 88 percent utilised supervised ML models (e.g., decision trees, support vector machines, random forests, and logistic regression) to forecast student performance. Their findings highlight the dominance of predictive modelling over interpretative analysis, while calling for greater methodological diversity and more substantial links to pedagogical theory.

Recent reviews on predictive learning analytics (PLA) confirm the increasing importance of ML in predicting academic success. Sghir et al. (2022) summarized a decade of PLA research, identifying a wide range of predicted outcomes (grades, retention, engagement) and data types (behavioural logs, assessment scores, demographic attributes). Despite significant progress, they pointed out a key gap: most models favour predictive accuracy over explainability, limiting their usefulness for educational decision-making. Alalawi et al. (2024) addressed this challenge by proposing a broader framework that combines pedagogical approaches with ML-based performance prediction and targeted intervention. Their Student Performance Prediction and Action (SPPA) model showed that integrating predictive analytics with instructor-led interventions can considerably enhance retention and pass rates, providing a replicable infrastructure for educators without needing institution-wide systems.

A parallel line of inquiry has examined the relationship between artificial intelligence and multimodal learning analytics. Mohammadi et al. (2025) reviewed 43 studies on AI-enhanced multimodal LA, concluding that while AI facilitates real-time feedback and customised learning, integration with learning theory and ethical governance remains limited. They identified gaps in deep AI implementation, data integration, and scalability, emphasising the need for interpretable and ethically aligned AI systems capable of managing complex, multimodal student data. These insights align with the broader educational movement towards transparency and fairness in AI applications.

Beyond higher education, machine learning has proven transformative for organizational and policy-level decision-making in development and quality assurance. Garbero et al. (2024) demonstrated how ML can be used to extract insights from large sets of project documentation, combining text mining and predictive analytics to strengthen data-driven evaluation frameworks. Their work parallels educational quality assurance systems, where diverse qualitative and quantitative indicators must be synthesized for informed decision-making. In the higher-education context, explainable AI (XAI) can similarly support continuous improvement cycles by linking institutional actions to measurable learning outcomes.

Complementary evidence links LA-driven analytics to student behavior, motivation, and well-being, which are variables that form the psychosocial foundation of learning. Walck-Shannon et al. (2021) found that the quality of learning strategies outweighs total study time, with active methods such as self-testing and elaboration predicting

higher exam performance. Aljaffer et al. (2024) reported that motivation, visual learning preferences, and memory recall significantly correlate with academic achievement among medical students. Lifestyle studies further reveal that balanced sleep and nutrition positively affect GPA (Mahfouz et al., 2024; Shafie et al., 2022), while irregular habits increase anxiety and stress (Rahimi et al., 2024; AlHamlan et al., 2025). Physical activity has been consistently associated with reduced stress and improved concentration, though standardized intervention protocols remain scarce (Guerriero et al., 2025).

Psychological health is a key factor in student success. Pascoe et al. (2019) established that academic strain directly harms well-being and learning outcomes. More recent research confirms that high stress and poor sleep quality increase the risks of anxiety and depression (Rahimi et al., 2024), while regular exercise helps reduce these effects. These behavioural insights highlight the importance of including psychological and physiological factors in predictive and explanatory models of academic performance. De Filippis and Foysal (2024) showed that anxiety and sleep quality are among the most important predictors of student stress, emphasising the need for comprehensive analytical frameworks that consider cognitive, emotional, and environmental factors.

The COVID-19 pandemic further highlighted how these factors are interconnected. Studies by Salazar-Granizo et al. (2024a, 2024b) showed that extended online learning disrupted healthy routines and increased stress levels, while face-to-face settings supported well-being. Pérez-Jorge et al. (2025) found that overload and imbalance of time were major stressors, but structured planning and emotional support helped lessen negative effects. These findings have led to international calls for universities to develop resilience-focused quality strategies that include both technological innovation and student well-being monitoring.

Taken together, the literature shows a clear shift from data collection and performance prediction towards interpretability, ethics, and holistic educational quality. Contemporary frameworks such as Rodríguez-Ortiz et al. (2025) and Mohammadi et al. (2025) agree on the importance of transparent and explainable AI to ensure that analytics improve, rather than hinder, human understanding of learning. This reflects a broader change in educational research from algorithmic optimisation to making meaning and gaining actionable insights. In this context, explainable ML techniques like SHapley Additive exPlanations (SHAP) are an important methodological step, enabling educators to trace how individual behaviours and environmental factors influence learning outcomes.
Despite these advances, most earlier studies treat prediction and explanation as separate aims. Few have systematically modelled behavioural, psychological, and environmental factors together within understandable frameworks. This study addresses this gap by using SHAP-based analysis of large-scale student data to show how daily habits, motivation, stress, sleep, and study environment together influence academic success. By combining interpretable ML methods with institutional quality goals, this research builds on previous work in learning analytics towards a clearer, more human-centred approach that directly supports the aims of United Nations Sustainable Development Goal 4 (Quality Education).

3. Methodology

3.1 Dataset and Preprocessing

The dataset used in this study is the *Student Habits and Academic Performance* dataset (Kumar, 2025). It contains data on 80,000 students' demographic characteristics, lifestyle habits, study environments, and academic performance. The dataset includes both numerical and categorical features, with "GPA" serving as the target variable representing students' academic success.

The features include demographic features (age, gender, family_income_range, parental_education_level), lifestyle and behavioral features (study_hours_per_day, social_media_hours, netflix_hours, sleep_hours, diet_quality, exercise_frequency, screen_time), academic features (semester, major, attendance_percentage, access_to_tutoring, study_environment, time_management_score, learning_style), psychological and emotional features (stress_level, mental_health_rating, exam_anxiety_score, motivation_level), and social and environmental features (part_time_job, extracurricular_participation, social_activity, parental_support_level, internet_quality). This categorization ensures that both internal and external factors influencing students' academic outcomes are comprehensively represented, covering demographic background, daily habits, cognitive and emotional states, academic behaviors, and social conditions.

Prior to analysis, the dataset was carefully cleaned and preprocessed. All rows containing missing values were removed to ensure data quality and model reliability. Non-numerical variables were converted into numerical form through logical enumerations. For instance, the *diet_quality* variable was encoded as Poor (0), Fair (1), and Good (2) to represent increasing levels of nutritional quality, while *study_environment* was ordered from less quiet to more quiet, Café (0), Dorm (1), Co-Learning Group (2), Library (3), and Quiet Room (4). This ordinal encoding

approach allows the model and subsequent SHAP analysis to capture meaningful progressions in the data rather than arbitrary categorical values.

All categorical enumerations applied during preprocessing are summarized in Table 1. Finally, all features were reviewed for consistency, ensuring that the resulting dataset was clean, standardized, and suitable for supervised regression modeling and SHAP-based interpretability analysis.

Table 1. Enumeration map of non-numerical features

Column Name	Category	Encoded As
gender	Male	0
	Female	1
	Other	2
major	Arts	0
	Business	1
	Psychology	2
	Biology	3
	Computer Science	4
	Engineering	5
part_time_job	No	0
	Yes	1
diet_quality	Poor	0
	Fair	1
	Good	2
parental_education_level	High School	0
	Some College	1
	Bachelor	2
	Master	3
	PhD	4
internet_quality	Low	0
	Medium	1
	High	2
extracurricular_participation	No	0
	Yes	1
study_environment	Cafe	0
	Dorm	1
	Co-Learning Group	2
	Library	3
	Quiet Room	4
access_to_tutoring	No	0
	Yes	1
family_income_range	Low	0
	Medium	1
	High	2
learning_style	Auditory	0
	Reading	1
	Visual	2
	Kinesthetic	3

3.2. Machine Learning and Explainability

To predict students' academic success, as measured by GPA, several supervised regression models were employed. Specifically, Random Forest, Logistic Regression, and Naïve Bayes algorithms were implemented to capture both linear and non-linear relationships between student characteristics and academic outcomes. Each model was trained using the preprocessed dataset described earlier, and performance was evaluated through mean absolute error (MAE) and mean absolute percentage error (MAPE).

After model comparison, the highest-performing model was selected for explainability analysis. To ensure transparency and interpretability of the predictions, SHAP was applied to this model. SHAP values quantify each feature's contribution to the predicted GPA, enabling identification of the factors that most strongly influence higher or lower academic performance.

The detailed results of the regression models and SHAP-based feature explanations are presented in the following section.

4. Results

4.1. Machine Learning Results

The performance of the three regression models, Random Forest, Logistic Regression, and Naïve Bayes, was evaluated and compared based on two key metrics: Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). The results of this comparison are summarized in Table 2.

Table 2. Comparison of machine learning models' performances

	MAE	MAPE
Random Forest	0.3451	10.28%
Logistic Regression	0.4211	13.21%
Naïve Bayes	0.6452	20.27%

Among the tested models, the Random Forest regressor achieved the highest predictive accuracy, demonstrating superior ability to capture complex, non-linear relationships in the data. Specifically, it attained an MAE of 0.3451 and a MAPE of 10.28%, outperforming the other models by a clear margin.

Given its strong predictive performance and robustness, the Random Forest model was selected as the final model for the subsequent SHAP- and ICE-based explainability analysis, which is presented in the next subsection.

4.2. SHAP and ICE Explainability Results

To interpret the contribution of each feature to the predicted student GPA, SHAP analysis was applied to the Random Forest model. Two complementary visualizations were generated: the SHAP beeswarm plot (Figure 1, left) and the SHAP bar plot (Figure 1, right). Together, these plots reveal which features most strongly influence academic success and how their values affect GPA predictions.

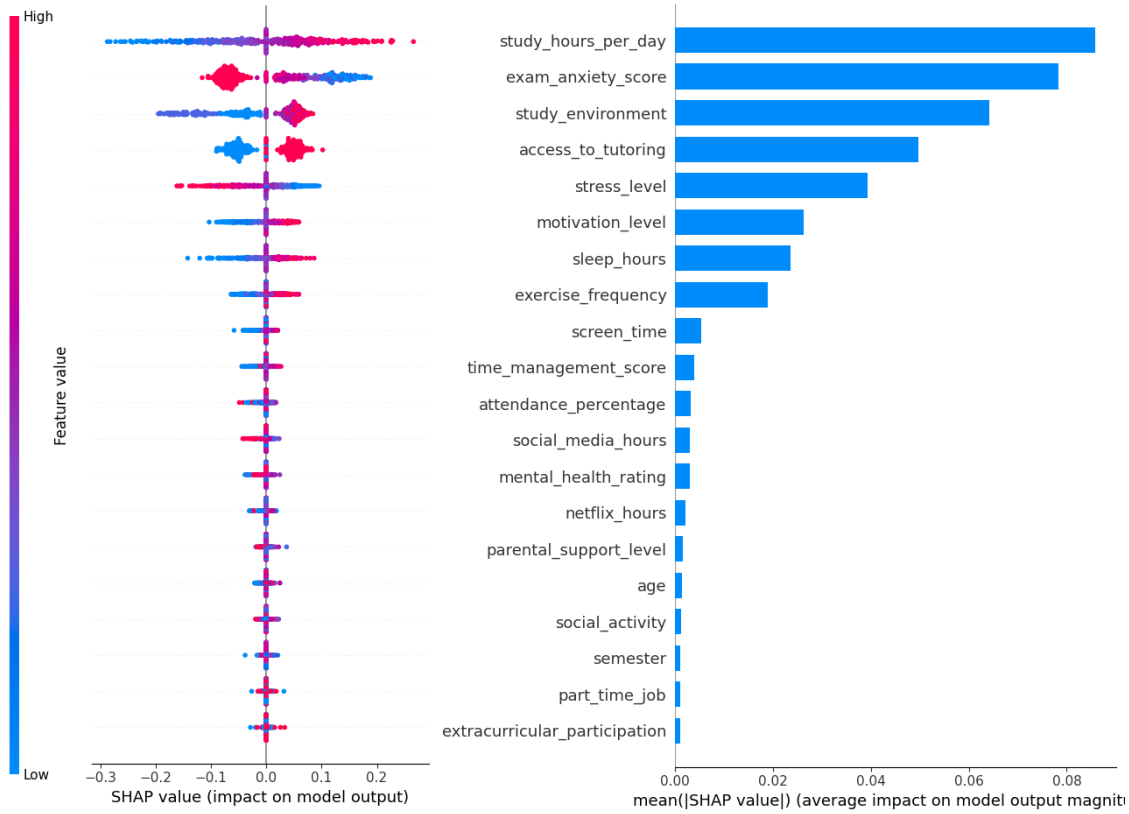


Figure 1. The SHAP Beeswarm (left) and Bar (right) plots

The beeswarm plot (Figure 1, left) provides a detailed, instance-level view of feature importance and direction of influence. The y-axis lists the features ranked by their overall importance, while the x-axis shows the SHAP value, representing each feature's impact on the model's output (i.e., predicted GPA). Dots to the right of the vertical line (positive SHAP values) indicate contributions that increase the predicted GPA, whereas dots to the left (negative SHAP values) indicate contributions that decrease it.

Each dot corresponds to a single student, and the color represents the feature value, pink for high values and blue for low values. For example, in the study_hours_per_day feature, pink dots appear predominantly on the right side, meaning that higher study hours are associated with higher GPAs. Conversely, for exam_anxiety_score, pink dots are concentrated on the left, showing that higher anxiety levels tend to lower GPA predictions.

The spread of dots along the x-axis indicates the magnitude of variation in that feature's effect on GPA. Features such as study_hours_per_day, exam_anxiety_score, and study_environment show a wide spread, indicating strong and varied influences on academic performance across students. In contrast, features near the bottom, such as part_time_job or extracurricular_participation, have narrow distributions, implying limited impact on GPA.

The encoded values used for categorical variables (e.g., study environment encoded from 0 = Café to 4 = Quiet Room) enable this interpretation. Because the categories were ordered logically, higher encodings correspond to conditions generally more conducive to learning. This ordering makes the SHAP patterns meaningful. For instance, higher study_environment values (quieter settings) are associated with higher GPA predictions.

The SHAP bar plot (figure 1-right) aggregates the mean absolute SHAP values across all students to provide a global ranking of feature importance. The features are sorted from most to least influential based on their average impact on the model output.

According to the results, study_hours_per_day is the most important predictor of GPA, followed closely by exam_anxiety_score, study_environment, and access_to_tutoring. Students who dedicate more time to studying and have access to tutoring services tend to achieve higher academic performance, while those with higher anxiety levels generally perform worse. Stress_level, motivation_level, and sleep_hours also emerge as key determinants, highlighting the role of psychological and lifestyle factors in shaping academic success.

In contrast, features like part_time_job, semester, and extracurricular_participation exhibit relatively minor influence, suggesting that these external or temporal factors have limited direct effect on GPA compared to study habits and mental well-being.

To better understand how specific features affect student success, SHAP dependence plots and Individual Conditional Expectation (ICE) plots were generated for the top eight influential variables. Although the marginal contribution of any single feature, as reflected in the SHAP values, may appear modest (typically on the order of a 0.1 increase or decrease in predicted GPA) it should be emphasized that academic performance is shaped by the combined influence of multiple factors. When favorable (or unfavorable) conditions across several features align simultaneously, their cumulative effects can produce substantial shifts in predicted GPA, underscoring the multifaceted and interdependent nature of the determinants of student success.

The SHAP dependence plots illustrate how the actual value of each feature (x-axis) relates to its SHAP value (y-axis), which represents its positive or negative contribution to the predicted GPA. Each dot in these plots corresponds to an individual student. When SHAP values increase with the feature value, the relationship is positive, meaning that higher values of that feature contribute to a higher GPA. The ICE plots complement this analysis by showing the predicted GPA (y-axis) as a function of the feature value (x-axis) for individual cases, represented by thin lines, with a bold blue line indicating the average ICE across all instances. These plots reveal both the average effect and the heterogeneity in how changes in the feature influence predictions across students. In the dependence plot for study_hours_per_day (Figure 2-a), a clear positive trend is visible. As daily study hours increase, SHAP values also rise, indicating that longer study durations contribute positively to GPA predictions. Students who study fewer than approximately 4 hours per day generally have negative SHAP values, suggesting a negative effect on predicted GPA, while those who study more than 4 hours per day show a consistent positive contribution. This pattern highlights the strong influence of regular study habits on academic success. The corresponding ICE plot (Figure 2-b) reinforces this trend, with the average ICE line increasing steadily from approximately 3.3 to 3.7 as study hours rise from 0 to 7, demonstrating that higher study time generally leads to higher predicted GPAs. Individual lines exhibit some variation, but the overall upward trend indicates a consistent positive impact across most students.

In Figure 2-c, the SHAP dependence plot for exam_anxiety_score shows a clear negative relationship between exam anxiety and predicted GPA. As the anxiety score increases along the x-axis, the SHAP values decrease, indicating that higher anxiety levels contribute negatively to academic performance. Students with lower anxiety scores (around 5-6) have distinctly positive SHAP values, meaning their exam calmness positively influences GPA predictions. In contrast, when anxiety levels rise above 8, SHAP values drop below zero, reflecting a detrimental effect on predicted GPA. This pattern suggests that excessive exam anxiety undermines students' academic outcomes, while maintaining moderate or low anxiety supports better performance. The ICE plot in Figure 2-d further illustrates this, with the average ICE line declining from about 3.7 to 3.3 as anxiety scores increase from 5 to 10. The spread of individual lines highlights heterogeneity, where some students experience steeper declines in predicted GPA with rising anxiety, emphasizing the variable sensitivity to this factor.

In Figure 2-e, the SHAP dependence plot for study_environment reveals a non-linear trend, indicating that quieter and more structured study settings are generally linked to higher predicted GPAs. The x-axis represents the ordered study environments-Café (0), Dorm (1), Co-Learning Group (2), Library (3), and Quiet Room (4), while the y-axis indicates their contribution to GPA predictions. Students studying in cafés and especially dorms tend to have negative SHAP values, suggesting these environments negatively impact academic performance. The negative influence is stronger for dorms than cafés, likely due to the higher levels of distraction and lack of dedicated study space typically found in dorm rooms. In contrast, environments such as co-learning groups, libraries, and quiet rooms show less negative or positive SHAP values, indicating that structured and distraction-free spaces enhance learning effectiveness and contribute to higher GPAs. The ICE plot in Figure 2-f complements this by depicting the average ICE line starting around 3.5 in café environments, dipping to approximately 3.4 in dorms, and then rising to about 3.6 for quieter settings. Individual lines exhibit variation, with some showing pronounced dips in predicted GPA for less optimal environments, underscoring the context-dependent effects on student outcomes.

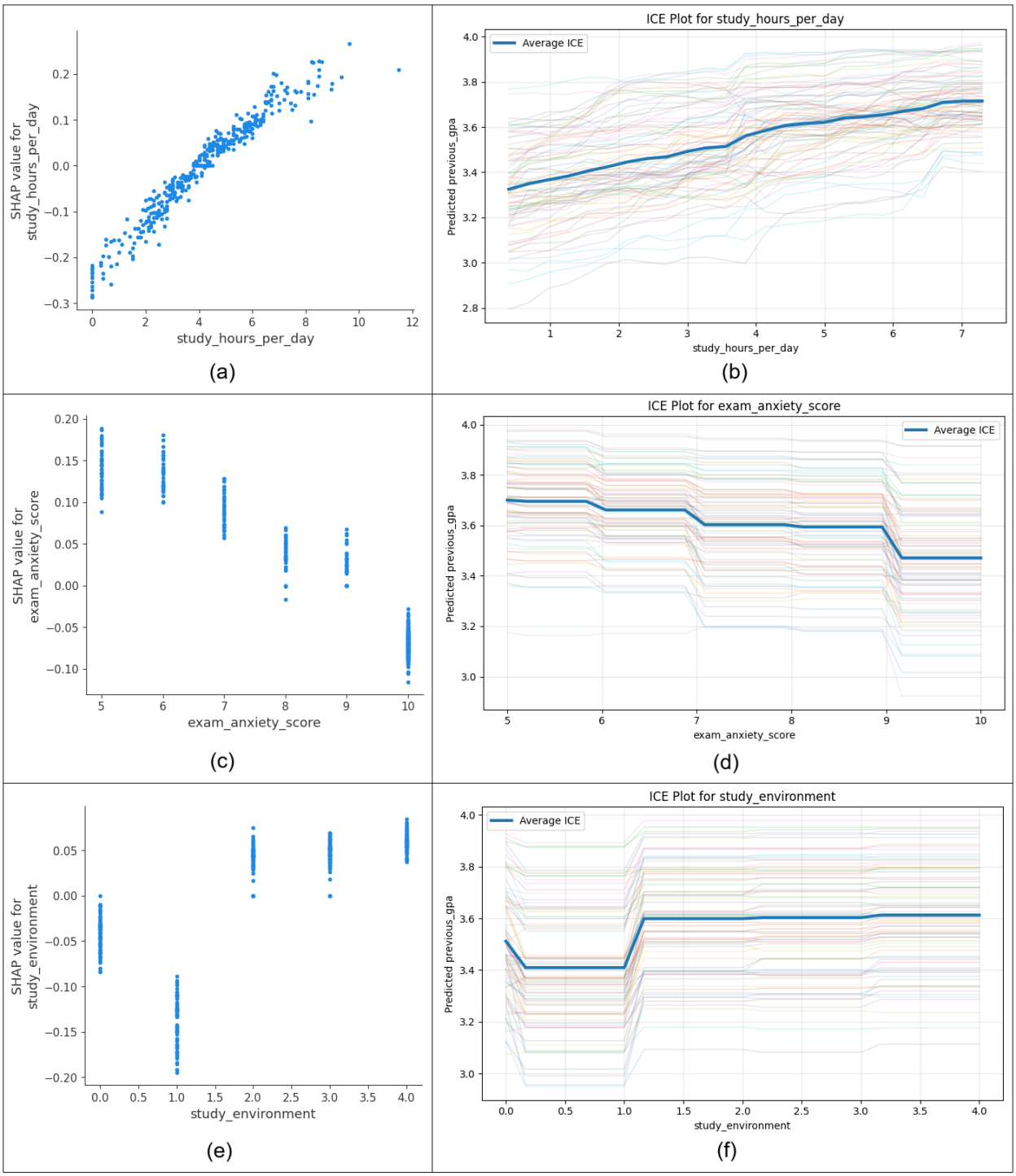


Figure 2. SHAP dependence and ICE plots for study hours per day (a, b), exam anxiety score (c, d), and study environment (e, f)

In Figure 3-a, the SHAP dependence plot for access_to_tutoring highlights a clear distinction between students with and without tutoring support. The feature is binary, where 0 represents no access to tutoring, and 1 indicates access to tutoring services. The plot shows that students without access to tutoring generally have negative SHAP values, meaning this condition contributes to lower predicted GPAs. Conversely, those with access to tutoring exhibit positive SHAP values, reflecting a significant positive effect on academic performance. This pattern underscores the importance of academic support mechanisms. Students who engage in tutoring or supplemental instruction are more likely to overcome learning challenges and achieve higher GPAs. The ICE plot in Figure 3-b supports this binary effect, with the average ICE line jumping from approximately 3.5 at no access to 3.6 with access. Individual lines cluster around these levels, indicating a relatively uniform positive shift in predicted GPA when tutoring is available.

In Figure 3-c, the SHAP dependence plot for stress_level shows a strong negative relationship between stress and predicted GPA. As stress levels increase along the x-axis, SHAP values steadily decline, indicating that higher stress contributes to lower academic performance. Students with stress levels below 4 generally have positive SHAP values, suggesting a beneficial or at least neutral influence on GPA. However, beyond this threshold, particularly at stress levels above 6, SHAP values drop sharply into the negative range. This pattern demonstrates that while moderate stress may have minimal impact, excessive stress severely hinders students' ability to perform well academically. The ICE plot in Figure 3-d confirms the negative trend, with the average ICE line decreasing from about 3.6 to 3.5 as stress rises from 0 to 10. The individual lines vary in slope, revealing that some students are more adversely affected by increasing stress than others.

In Figure 3-e, the SHAP dependence plot for motivation_level displays a clear positive association with student success. As motivation increases along the x-axis, SHAP values rise steadily, indicating that higher motivation consistently contributes to better academic outcomes. Students with motivation levels below 4 tend to have negative SHAP values, indicating a detrimental effect on GPA, whereas those scoring above 6 exhibit increasingly positive SHAP values. This upward trend suggests that maintaining high motivation is one of the strongest psychological factors supporting academic success, as motivated students are more likely to engage effectively in learning and sustain productive study behaviors. The ICE plot in Figure 3-f illustrates this positivity, with the average ICE line ascending from approximately 3.5 to 3.6 as motivation increases from 0 to 10. Individual lines show parallel increases, though with some dispersion, highlighting consistent but varying degrees of benefit across students.

In Figure 4-a, the SHAP dependence plot for sleep_hours reveals a positive, non-linear relationship between sleep duration and predicted GPA. As sleep hours increase from 4 to around 7-8 hours, SHAP values rise sharply, indicating that adequate sleep contributes significantly to better academic performance. Students sleeping fewer than 6 hours tend to have negative SHAP values, suggesting that sleep deprivation negatively affects GPA predictions. Beyond approximately 8 hours, the SHAP values plateau, implying that while sufficient rest is beneficial, excessive sleep does not provide additional academic advantage. This trend highlights that maintaining a balanced sleep schedule (around 7 to 8 hours per night) optimally supports learning and cognitive performance. The ICE plot in Figure 4-b depicts this non-linearity, with the average ICE line rising from about 3.5 at low sleep hours to a plateau around 3.6 at 7-9 hours. Individual lines follow similar patterns but with variations in the point of plateau, indicating personalized optimal sleep ranges.

In Figure 4-c, the SHAP dependence plot for exercise_frequency shows a clear positive relationship between physical activity and predicted GPA. As the number of exercise days per week increases along the x-axis, SHAP values gradually rise, indicating that more frequent exercise contributes positively to academic performance. Students who exercise fewer than two times per week generally have negative SHAP values, reflecting lower predicted GPAs, while those engaging in regular exercise (five or more times per week) consistently show positive SHAP contributions. This pattern suggests that maintaining an active lifestyle not only supports physical health but also enhances cognitive functioning, concentration, and overall academic success. The ICE plot in Figure 4-d reinforces this, with the average ICE line increasing from approximately 3.5 to 3.6 as exercise frequency rises from 0 to 7. The spread of individual lines suggests heterogeneity, where the positive effect is more pronounced for some students than others.

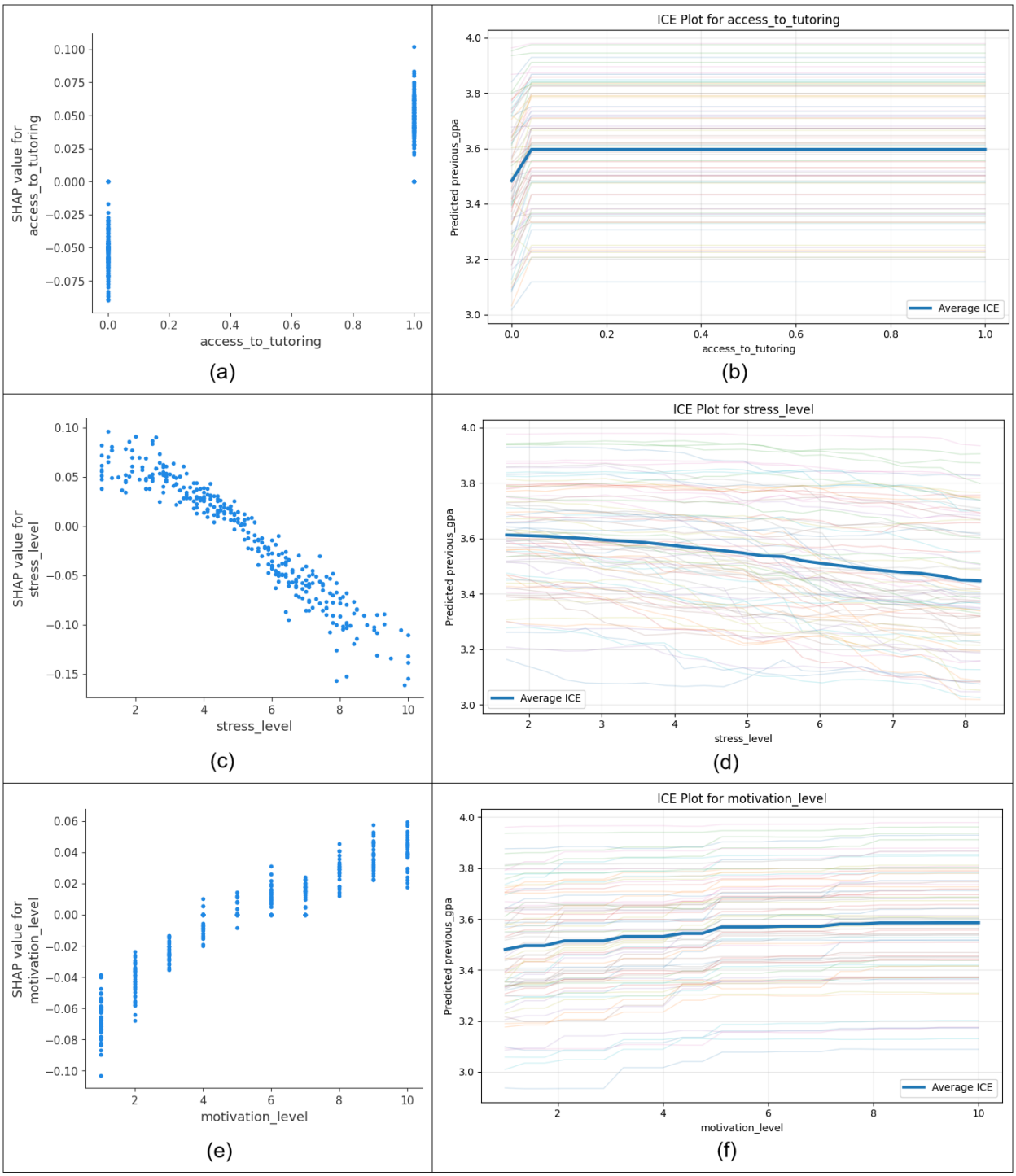


Figure 3. SHAP dependence and ICE plots for access to tutoring (a, b), stress level (c, d), and motivation level (e, f)

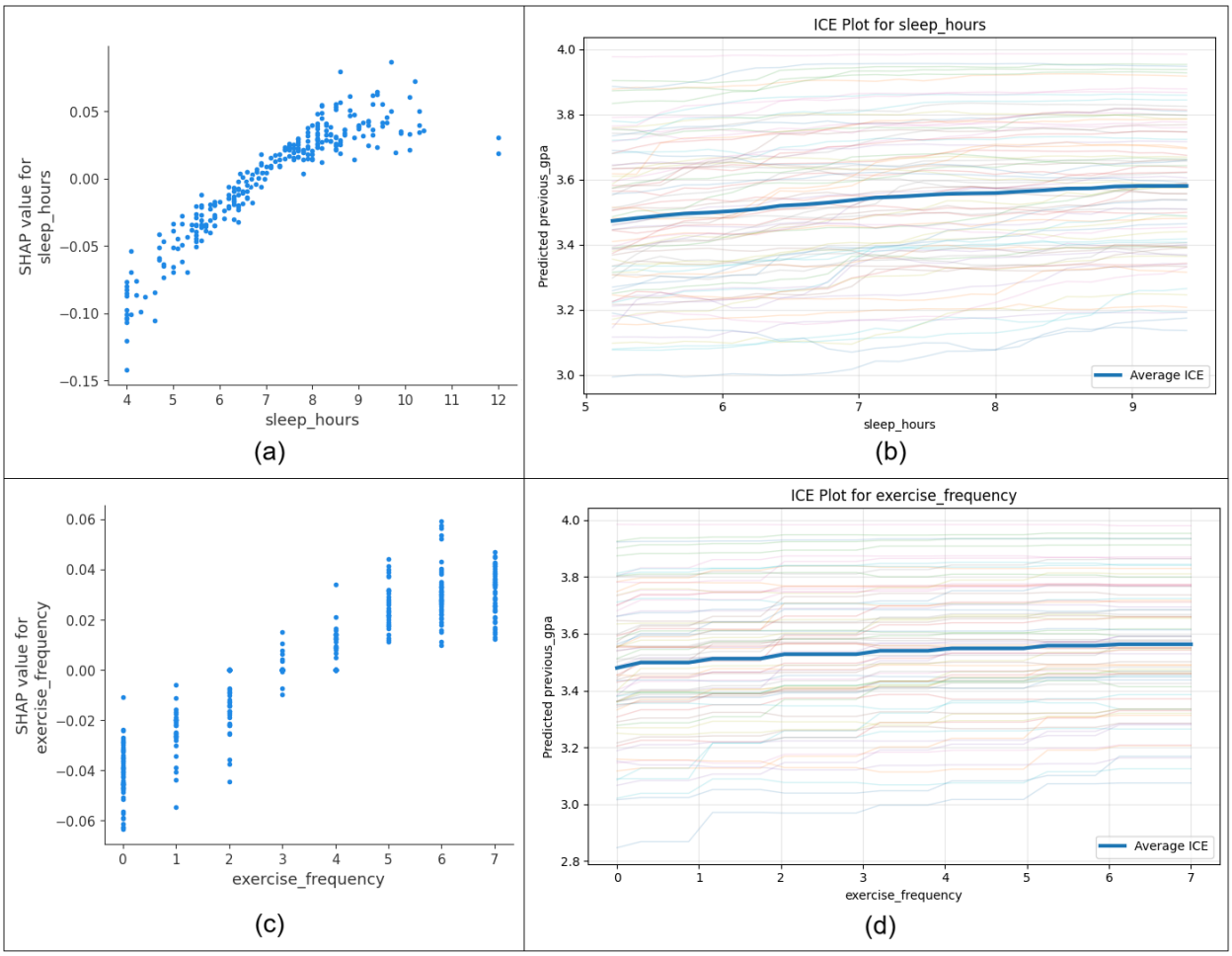


Figure 4. SHAP dependence and ICE plots for sleep hours (a, b) and exercise frequency (c, d)

To further elucidate the local interpretability of the model, SHAP waterfall plots were employed to examine the feature contributions for three randomly selected students. These plots demonstrate how each feature's SHAP value modifies the base expected GPA of 3.604, resulting in the final predicted GPA for the individual student. Positive contributions are shown in blue and increase the prediction, whereas negative contributions are shown in red and decrease it. The analysis begins with a student whose features collectively elevate the predicted GPA above the baseline (Figure 5), proceeds to a student in whom positive and negative contributions largely offset one another, yielding a predicted GPA near the expected value (Figure 6), and concludes with a student whose features predominantly reduce the predicted GPA below the baseline (Figure 7).

In Figure 5, the waterfall plot illustrates a case where the predicted GPA is elevated to 3.936. The most influential positive drivers are a moderate exam anxiety score and a relatively high number of daily study hours, both of which contribute substantially to raising the prediction above the baseline. These effects are further supported by adequate sleep duration, very high motivation, a reasonably favorable study environment (Co-Learning Group), frequent exercise, good mental health, and limited social media use. The remaining features have only negligible or neutral influence. This configuration demonstrates that a balanced psychological state, characterized by manageable anxiety levels, combined with strong study discipline, sufficient rest, intrinsic motivation, and healthy lifestyle habit can collectively produce a meaningful improvement in predicted academic performance.

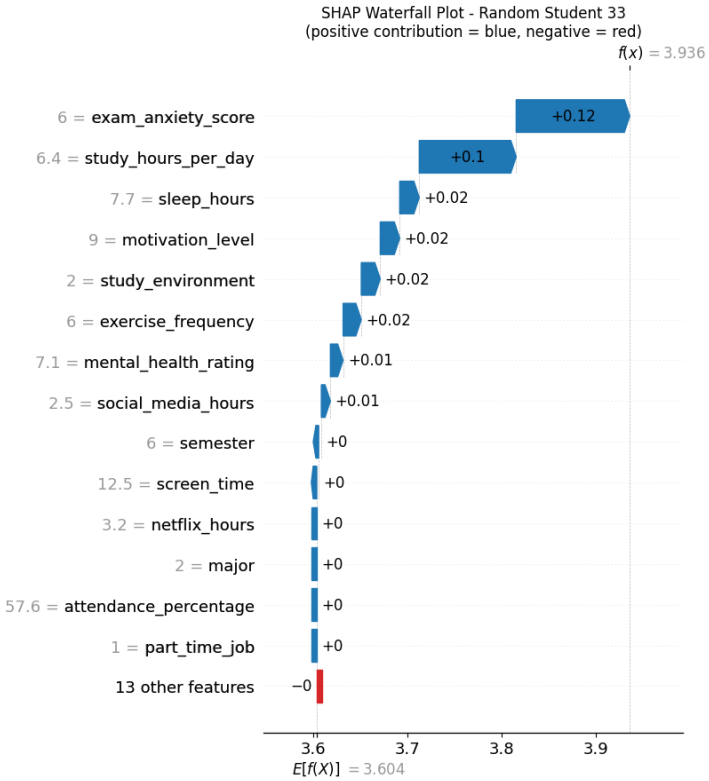


Figure 5. SHAP Waterfall Plot for a Random Student with Features Pushing Predicted GPA Higher (Student 33).

In Figure 6, the waterfall plot depicts a student whose predicted GPA remains close to the expected value at 3.642, reflecting a near-neutral net effect. The prediction is bolstered by a solid amount of daily study time, an optimal study environment (Quiet Room), moderate stress (which in this range appears to exert a mildly positive influence), and sufficient sleep. Smaller positive contributions arise from age, major, time management ability, attendance, diet quality, and limited recreational screen time. However, these gains are largely counteracted by very high exam anxiety, extremely low motivation, and lack of access to tutoring. This example highlights how strengths in behavioral and contextual factors can be substantially undermined by intense psychological distress and insufficient motivational drive or academic support, resulting in an outcome that deviates only minimally from the average prediction.

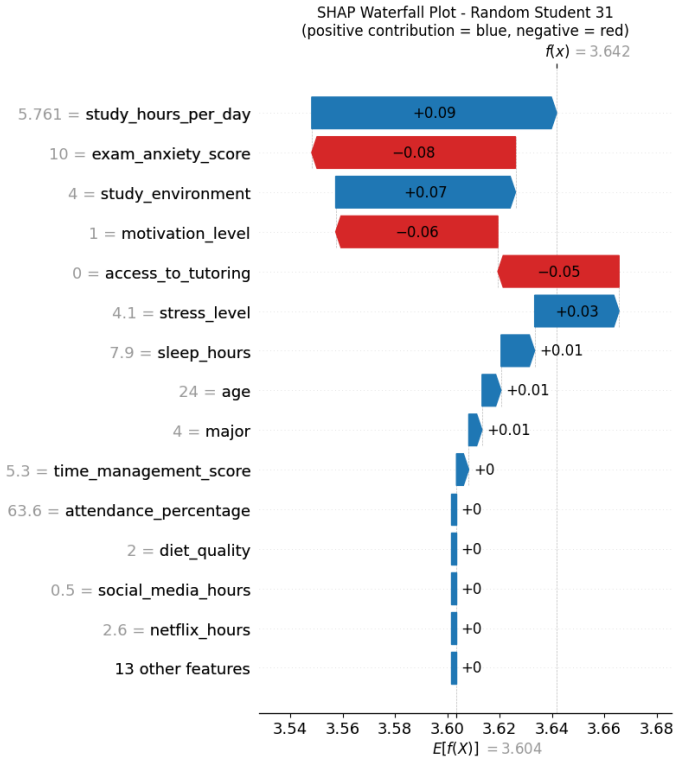


Figure 6. SHAP Waterfall Plot for a Random Student with Features Largely Canceling Each Other (Student 31).

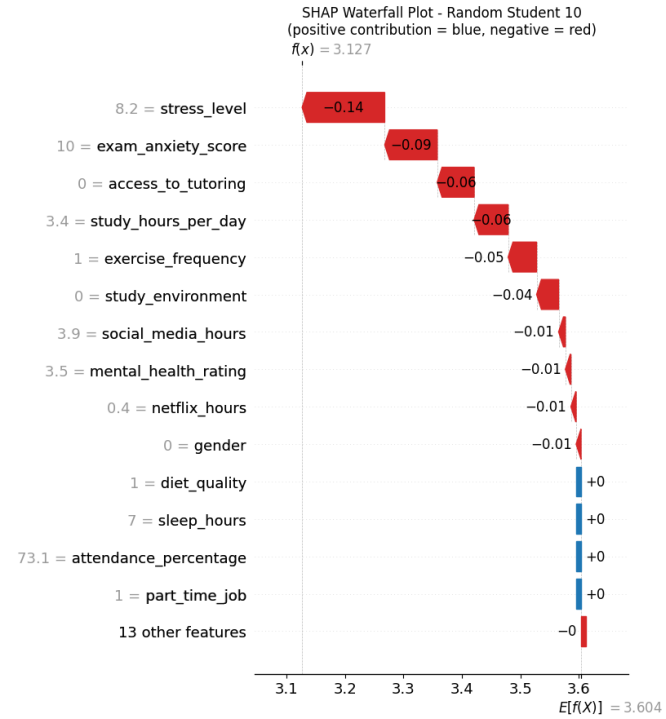


Figure 7. SHAP Waterfall Plot for a Random Student with Features Pushing Predicted GPA Lower (Student 10).

In Figure 7, the waterfall plot reveals a student whose predicted GPA is reduced to 3.127. The strongest negative influences stem from very high stress levels and maximum exam anxiety, both of which exert pronounced downward pressure on the prediction. Additional detrimental effects arise from insufficient daily study time, infrequent exercise, a highly distracting study environment (Café), excessive social media use, poorer mental health, and suboptimal diet quality. The absence of tutoring access further compounds the negative impact. Features such as sleep duration, attendance, part-time employment status, and most remaining variables contribute negligibly or neutrally and fail to offset the cumulative adverse effects. This profile underscores the severe

consequences of elevated psychological strain combined with inadequate study habits, poor environmental conditions, limited physical activity, and lack of academic support mechanisms.

5. Discussion

The findings of this study offer actionable insights for advancing Quality Education (UN SDG 4) through evidence-based decision-making in higher education. By integrating interpretable machine learning with both global (SHAP beeswarm/bar/dependence and ICE plots) and local (SHAP waterfall) explainability techniques, the research illustrates how large-scale student data can be transformed into transparent and meaningful indicators of academic well-being and performance. The results consistently highlight the strong influence of behavioral, psychological, and environmental factors (most notably study hours, motivation level, stress, sleep duration, study environment, and access to tutoring) on GPA outcomes. These patterns should be carefully considered by higher education institutions and policy-makers seeking to strengthen learning quality, student support systems, and academic equity.

From an institutional perspective, the interpretable patterns revealed suggest the value of developing data-informed student success frameworks. Universities can implement personalized advising systems capable of identifying students at risk due to insufficient study engagement, elevated stress or anxiety, suboptimal sleep, or limited access to support services, and then guide them toward targeted interventions such as structured study programs, counseling, tutoring, or improved study spaces. For instance, the observed thresholds, such as markedly reduced predicted performance below four hours of daily study or in highly distracting environments, underscore the potential benefits of proactive measures to enhance study conditions and well-being resources.

At the policy level, the insights call for integrating learning analytics and well-being monitoring into institutional and national quality assurance frameworks. Explainable AI tools, including both aggregate-level dependence/ICE visualizations and instance-specific explanations, can support continuous evaluation of student learning conditions, mental health indicators, and resource accessibility. Funding priorities could accordingly be directed toward expanding tutoring availability, mental health services, and initiatives that promote healthy lifestyle habits and motivation, all of which have been shown to exert measurable effects on academic outcomes.

Ultimately, the study reinforces that quality in higher education extends far beyond curriculum design or institutional metrics and encompasses the holistic support of the learner. The adoption of transparent, explainable AI approaches enables universities to design interventions that are equitable, evidence-based, and aligned with the human-centered objectives of SDG 4, thereby nurturing inclusive and high-quality education capable of empowering all students to achieve their potential.

6. Conclusion

The results of this study demonstrate that interpretable machine learning, augmented by SHAP dependence plots, ICE plots, and local SHAP waterfall explanations, can yield deep and actionable understanding of the determinants of student success. It has been shown that academic performance is significantly shaped by a combination of behavioral factors (study hours, sleep duration, exercise frequency), psychological dimensions (motivation level, exam anxiety, stress, mental health), and environmental conditions (study environment, access to tutoring), rather than by cognitive ability alone. These influences can be monitored and targeted through an institutional intervention.

From a quality standpoint, the presented framework offers a practical and transparent pathway toward realizing UN SDG 4, Quality Education. By employing explainable AI techniques at both global and individual levels, universities can monitor learning conditions, identify at-risk students early, optimize support services, and cultivate healthier academic environments in a manner that is fair, interpretable, and student-centered. Such approaches ensure that data-driven policies remain accountable and aligned with educational equity goals.

In conclusion, integrating explainable artificial intelligence into learning analytics represents a significant advancement in educational quality management. When applied ethically and responsibly, these methods empower decision-makers to convert complex student data into targeted, evidence-informed actions that enhance learning outcomes. Accordingly, this study contributes both methodological rigor and strategic guidance to the continuing efforts of higher education institutions and policy-makers to deliver inclusive, equitable, and high-quality educational experiences in accordance with the principles of SDG 4.

Note: This article is an extended and revised version of the paper titled "Quality-Centered Learning Analytics: Interpretable Modeling of Students' Success with Explainable AI," presented at the International Conference on

Quality in Higher Education in November 2025. The present work incorporates substantial new content, additional analyses, updated sections, and minor changes to the title and structure.

References

Alalawi, K., Athauda, R., & Chiong, R. (2024). An Extended Learning Analytics Framework Integrating Machine Learning and Pedagogical Approaches for Student Performance Prediction and Intervention. International Journal of Artificial Intelligence in Education, 35(3), 1239–1287. https://doi.org/10.1007/s40593-024-00429-7

AlHamlan, A. W., AlDhafiri, D. A., AlHajri, N. A., AlHouli, A. H., AlHubaidah, S. M., AlJassar, A. M., AlMutairi, N. N., AlRabiah, A. S., Al-Sultan, A. T., & Qasem, W. (2025). The impact of academic stress on the lifestyle of university students in Kuwait. BMC Public Health, 25(1). https://doi.org/10.1186/s12889-025-23569-4

Aljaffer, M. A., Almadani, A. H., AlDughaither, A. S., Basfar, A. A., AlGhadir, S. M., AlGhamdi, Y. A., AlHubaysh, B. N., AlMayouf, O. A., AlGhamdi, S. A., Ahmad, T., & Abdulghani, H. M. (2024). The impact of study habits and personal factors on the academic achievement performances of medical students. BMC Medical Education, 24(1). https://doi.org/10.1186/s12909-024-05889-y

de Filippis, R., & Foysal, A. A. (2024). Comprehensive analysis of stress factors affecting students: a machine learning approach. Discover Artificial Intelligence, 4(1). https://doi.org/10.1007/s44163-024-00169-6

Doroudi, S. (2024). On the paradigms of learning analytics: Machine learning meets epistemology. Computers and Education: Artificial Intelligence, 6, 100192. https://doi.org/10.1016/j.caeai.2023.100192

Ersozlu, Z., Taheri, S., & Koch, I. (2024). A review of machine learning methods used for educational data. Education and Information Technologies, 29(16), 22125–22145. https://doi.org/10.1007/s10639-024-12704-0

Garbero, A., Carneiro, B., Cerulli, G., & Resce, G. (2024). Leveraging artificial intelligence and machine learning analytics for development effectiveness: a comprehensive approach to extract insights from project data. Journal of Development Effectiveness, 17(1), 71–100. https://doi.org/10.1080/19439342.2024.2361008

Gasevic, D., Rose, C., Siemens, G., Wolff, A., & Zdrahal, Z. (2014). Learning analytics and machine learning. In Proceedings of the Fourth International Conference on Learning Analytics And Knowledge (pp. 287–288). LAK '14: Learning Analytics and Knowledge Conference 2014. ACM. https://doi.org/10.1145/2567574.2567633

Guerriero, M. A., Dipace, A., Monda, A., De Maria, A., Polito, R., Messina, G., Monda, M., di Padova, M., Basta, A., Ruberto, M., Capasso, E., Moscatelli, F., & Limone, P. (2025). Relationship Between Sedentary Lifestyle, Physical Activity and Stress in University Students and Their Life Habits: A Scoping Review with PRISMA Checklist (PRISMA-ScR). Brain Sciences, 15(1), 78. https://doi.org/10.3390/brainsci15010078

Kumar, A. (2025). *Student Habits and Academic Performance Dataset* [Dataset]. Kaggle. https://www.kaggle.com/datasets/aryan208/student-habits-and-academic-performance-dataset

López-Meneses, E., López-Catalán, L., Pelícano-Piris, N., & Mellado-Moreno, P. C. (2025). Artificial Intelligence in Educational Data Mining and Human-in-the-Loop Machine Learning and Machine Teaching: Analysis of Scientific Knowledge. Applied Sciences, 15(2), 772. https://doi.org/10.3390/app15020772

Mahfouz, H. A., Alhazmi, N. F., Almatrafi, M. K., Almehmadi, S. S., Alharbi, J. K., Qadi, L. R., & Tawakul, A. (2024). The Influence of Lifestyle on Academic Performance Among Health Profession Students at Umm Al-Qura University. Cureus. https://doi.org/10.7759/cureus.56759

Mohammadi, M., Tajik, E., Martinez-Maldonado, R., Sadiq, S., Tomaszewski, W., & Khosravi, H. (2025). Artificial intelligence in multimodal learning analytics: A systematic literature review. Computers and Education: Artificial Intelligence, 8, 100426. https://doi.org/10.1016/j.caeai.2025.100426

Pascoe, M. C., Hetrick, S. E., & Parker, A. G. (2019). The impact of stress on students in secondary school and higher education. International Journal of Adolescence and Youth, 25(1), 104–112. https://doi.org/10.1080/02673843.2019.1596823

Pérez-Jorge, D., Boutaba-Alehyan, M., González-Contreras, A. I., & Pérez-Pérez, I. (2025). Examining the effects of academic stress on student well-being in higher education. Humanities and Social Sciences Communications, 12(1). https://doi.org/10.1057/s41599-025-04698-y

Rahimi, A., Wardak, M. F., & Shayan, N. A. (2024). Assessing the relationship between lifestyle factors and mental health outcomes among Afghan university students. Journal of Affective Disorders Reports, 17, 100827. https://doi.org/10.1016/j.jadr.2024.100827

Rodríguez-Ortiz, M. Á., Santana-Mancilla, P. C., & Anido-Rifón, L. E. (2025). Machine Learning and Generative AI in Learning Analytics for Higher Education: A Systematic Review of Models, Trends, and Challenges. Applied Sciences, 15(15), 8679. https://doi.org/10.3390/app15158679

Salazar-Granizo, Y. E., Hueso-Montoro, C., & Caparros-Gonzalez, R. A. (2024-a). Lifestyles and Academic Stress in University Students of Health Sciences: A Mixed-Methodology Study. Healthcare, 12(14), 1384. https://doi.org/10.3390/healthcare12141384

Salazar-Granizo, Y. E., Hueso-Montoro, C., & Caparros-Gonzalez, R. A. (2024-b). Lifestyles and academic stress among health sciences students at the National University of Chimborazo, Ecuador: a longitudinal study. Frontiers in Public Health, 12. https://doi.org/10.3389/fpubh.2024.1447649

Sciarrone, F. (2018). Machine Learning and Learning Analytics: Integrating Data with Learning. In 2018 17th International Conference on Information Technology Based Higher Education and Training (ITHET) (pp. 1–5). 2018 17th International Conference on Information Technology Based Higher Education and Training (ITHET). IEEE. https://doi.org/10.1109/ithet.2018.8424780

Sghir, N., Adadi, A., & Lahmer, M. (2022). Recent advances in Predictive Learning Analytics: A decade systematic review (2012–2022). Education and Information Technologies, 28(7), 8299–8333. https://doi.org/10.1007/s10639-022-11536-0

Shafie, S. N. M., Shahri, M. S., Izuddi, N. N. N. I., Shukri, N. M., Aziz, N. A., Amran, A., & Nafi, M. N. A. (2022). Association between Lifestyle-Related Behaviors and Academic Performance among Students. In International Academic Symposium of Social Science 2022 (p. 105). International Academic Symposium of Social Science. MDPI. https://doi.org/10.3390/proceedings2022082105

Vinoth Kumar, E. S., Augustian Isaac, R., Sundaravadivel, P., & Janaki Ramal, P. (2025). Machine learning-driven development of a behaviour-based student classification system (SCS-B) for enhanced educational analytics. Scientific Reports, 15(1). https://doi.org/10.1038/s41598-025-21332-8

Walck-Shannon, E. M., Rowell, S. F., & Frey, R. F. (2021). To What Extent Do Study Habits Relate to Performance? CBE—Life Sciences Education, 20(1), ar6. https://doi.org/10.1187/cbe.20-05-0091

Investigating Historical Skills Through the Lens of Film: A Case Study on Historical Thinking Skills Acquired Through Student-Produced Videos

J. Alex Boggs, Ph.D.
Assistant Professor of History at Liberty University
ORCID: https://orcid.org/0000-0003-2753-382X

Abstract
Communication is a necessary skill for students to develop in the 21st century, particularly with the advancement of new and novel practices of communication, but has not reached broad saturation in core curricula. The history classroom is well-poised as a context to build rich historical skills, digital literacy, and media literacy as part of a broader adoption of video production. This single instrumental case study evaluates the historical skills developed through student-generated video in graduate history classrooms. The theoretical framework includes Cognitive Constructivism (Bruner, 1977). and Activity Theory (Engeström, 2001). Ten participants were recruited from one site across different courses. The findings from interviews, focus groups, journal reflections, and video artifacts included four key themes: framing, evaluation, synthesis, and technical video skills. Participants drew connections between their historical training and implemented it into the design of their videos. They modeled the historical process through the production of the videos. The findings blend findings from historical thinking skills as well as distinct phases of the video production process. Implementation of similar projects in graduate history classrooms could warrant honing of historical skills and development of video production skills. Further research is necessary to understand its impact on primary, secondary, and undergraduate education.
Keywords: student-generated video production, historical thinking skills, media literacy, digital literacy, multimodal composition, semiotics.

Overview

Video is an ever-present medium of communication, which has revolutionized the spread of information in the twenty-first century. Film and the arrangement of images in quick and sequential order have been a communication tool since the late nineteenth century, with the advancement of inventors like the Lumiere brothers (Barnouw, 1993). Movies, videos, and film have changed considerably in terms of production costs, distribution, accessibility of creation, and advancement of filmmaking technology. Video stands as a means of communication that rivals the written word, particularly on internet platforms like YouTube, Twitch, and other social media.

Incorporating video into classroom environments has predominantly been in the form of passive or even active viewing on the part of students. Particularly in the field of history, documentaries have long been a pedagogical tool to support learning and immerse students in other cultures, ideas, and to present them with narrative representations of the past (Husbye & Vander Zanden, 2015). Just viewing videos does not give students adequate skill in communicating through the medium of film (Lam et al, 2021; Snelson, 2018). Further investigations into the world of multimedia composition support initial investigations into the topic of filmmaking in the classroom (Huang & Xia, 2024; Lam et al., 2021; Xia, 2024)

Studies to date on student-generated video in the classroom have been conducted in the sciences and in English (Reyna & Meier, 2018; Snelson, 2018). Scholars have also attempted categorization of filmmaking practices for the classroom by defining taxonomies of video style (Arruabarrena et al., 2021; Xia, 2024). Thus far, limited research has been done on the development of historical skills or video skills through the development of videos in the history classroom. Additionally, there are few studies that tackle multimodal video production within graduate education. History is a discipline that is rich in media literacy skills and it is also features strong storytelling to guide the overall narratives produced from research (Gaddis, 2004; Lukacs, 2011). The goal of this study is to bridge the gap between historical thinking skills and video production to provide a framework to introduce multimodal video as an activity to stimulate historical thinking. Finding alternative avenues to present historical research will retain interest from students, while also providing them opportunities to engage in deeper critical thinking, content acquisition, and digital technical skills.

Theory and Literature:

Two theories guide this research – Cognitive Constructivism and Activity Theory. Cognitive Constructivism is a natural part of the process of developing knowledge, particularly as it has been expressed by Jerome Bruner (1966). For history, the development of knowledge happens in phases both in content acquisition and in the skill

development to further analyze historical content. Cognitive Constructivism consists of the development of new knowledge through the incorporation of existing knowledge into new knowledge (Dewey, 1910; Bruner, 1966, 1977). Bruner describes a spiral method of knowledge construction whereby learners observe a central phenomenon from different angles and repeatedly, as if they were observing it by walking around it down a spiral staircase (Bruner, 1977). In Cognitive Constructivism as a whole, there are representative building blocks of information that form new information, partially described by schema, and further elucidated by the process of discovery (Dewey, 1910; Piaget, 1977). When learners first interact with a topic, it is often clumsy, confused, and ill-defined, but as the process of discovery unfolds, it becomes more defined and clear to the point of abstraction. Bruner (1966, 2006) represents these steps through enactive, iconic, and symbolic expressions. These expressions align with the field of history, a field which Bruner (1977) would label as intuitive, where knowledge is not always constructed evenly and takes several attempts to build fuller pictures of the whole field by several observations of primary and secondary accounts.

Within the classroom, Cognitive Constructivism leverages the malleability of the construction of knowledge into tangible steps for the learner. The learner starts in a context where the problem is ill-defined, but the teacher makes it manageable through scaffolding. Teachers then organize content to reinforce to students, moving them through optimal sequencing of material, then to patterns of reinforcement (Bruner, 1977, 2006). The process by which the learner masters skills and content can be summarized by acquisition, transformation, and evaluation of content (Bruner, 1966; Stapleton & Stefaniak, 2019). In history, these skills are connected through understanding basic concepts, connecting those concepts into causal relationships, and further developing learning through the evaluation of other sources to connect ideas (Bruner, 1966; Monte-Sano, 2016; Stapleton & Stefaniak, 2019). For this study, Cognitive Constructivism guides the discussion of the learning process of video production.

The development of a product also bears with it a context of creation. Where Cognitive Constructivism provides scaffolding for the mental work happening during the process of video production, Activity Theory gives structure to the work of students, but it also exists within a broader context of viewership (Engeström, 1987, 2001). For any learning event, there are tools, subjects, and outcomes for any given activity (Engeström, 1987; Gagne, 1970). Beyond these three central items are the broader social context of division of labor, community, and rules (Engeström, 1987). For students within the context of making a video, they are navigating a world where the rules of the discipline of history and the rules of filmmaking apply an implicit guideline for what is considered good communication. The video is not developed within a backdrop that is just the student alone, but one where the teacher reviews, peers may view, and the knowledge is not always clear. Activity Theory, in the context of this study, provides a framework for the production of videos removed from the learning processes. Working in tandem, Cognitive Constructivism provides a mental framework centering on historical thinking and Activity Theory provides the necessary context of rules and tools for students developing film through the focus on video skills.

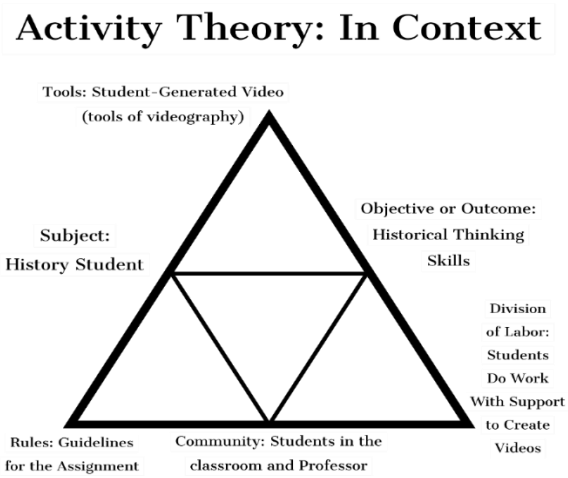


Figure 1.

Note. Model of Activity Theory adapted for use within the construct of this research study. Focuses on splitting out the different areas of the filmmaking process within the scope of Activity Theory. Adapted from Activity Theory Model in Wikimedia, n.d., (https://commons.wikimedia.org/w/index.php?curid=18600235). Copyright, 2012 by Matt Bury CC BY-SA 3.0.

As this topic combines several fields, digital literacy, historical skills, and video production, it is useful to understand the scope of the field as it stands today. Digital literacy is part of the process by which students demonstrate fluency with media available through the internet. Broadly, it is the manner in which students find, source, evaluate, interpret, consume, and construct digital media (Park et al., 2021). Engaging with different media types, perspectives of authors, or even varied levels of authenticity adds to the overall complexity of developing digital fluency (Jones-Jang et al., 2021). Social media represents a newer context in which students find videos (Masanet et al., 2019; Scolari, 2019). The mediating factor of algorithmic feeds providing users with direct information, vetted or unvetted, can lead to issues with understanding and making meaning of the content that is being consumed. If a person identifies with a figure online, the line between misleading and verified information can be blurred (Huang et al., 2023; Masanet et al., 2019; Scolari, 2019).

To be fully fluent within these media takes time to develop skills of understanding the grammar of the systems, let alone the language that is used (Guzmán-Simón et al., 2017; Karasavvvidis, 2019; Nichols & Stornaiuolo, 2019; Tiernan & Farren, 2017). For video specifically, skills include understanding the perspectives of the person speaking, the language of cuts and edits, visual representations, and semiotics, as well as putting these into action in their own constructions (Karasavvvidis, 2019; Mengis et al., 2018; Palmgren-Neuvonen & Korkeamaeki, 2015; Rainville & Gordh, 2016; Snelson, 2018). The construction of video material takes more than just the review and consumption of video material, but also includes strictly auditory clips, still photographs, and myriad written materials (Husbye & Vander Zanden, 2015; Park et al., 2021; Xia, 2024).

Digital literacy also includes the skills students need to use information available to them through AI sources. For the subject of video, AI represents a clunky but improving method of conversing through digital media (Orak & Turan, 2024). There are advancements that make AI a powerful tool for video creators, including storyboarding, image generation, and motion control enhancements, but also contain promises that are not fully realized yet (Huang et al., 2023). Tools for video production also contain some issues with privacy and ethics, including deepfakes or misleading videos (Huang et al., 2023). AI literacy currently overlaps the information literacy aspect of digital literacy in the impact it has on students in multimodal compositional contexts (Wu & Zhang, 2025).

This historical field has skills which overlap some of those found in the digital space, but are tuned for analyzing sources from the past. Seixas (2017) identified six key historical skills that students practice within the broader historical field: historical significance, primary source evidence, continuity and change, cause and consequence, historical perspective-taking, and moral reasoning. There are key primary source skills that form the core of analysis for students in the history classroom (DIG, 2020; Seixas, 2017). Primary source analysis forms the basis of historical thinking through the skills of sourcing, corroboration, contextualization, and close reading (DIG, 2020; Breakstone et al., 2018). Argumentation is the process by which students formulate or frame their evidence, claims, and warrants (Monte-Sano, 2016). Historical thinking revolves around the concept of meaning-making, where scholars and students must grapple with issues of time, contingency, and scale (Gaddis, 2004; Seixas, 2017). The softer skills of history are less evaluative in nature, but reflect more of the ethical or moral dimension of how learning history impacts students as it is learned (Carr, 1961; Fea, 2013; Seixas, 2017). While not a justification of actions, empathy can be a powerful tool to understand the motivations behind the actions of people in the past (Fea, 2013; Lowenthal, 1987). For the purposes of this study, these historical skills can be simplified to three core areas, framing, evaluation, and synthesis.

Video production includes its own distinct cycles: media review, pre-production, production, and post-production, and distribution (Bowen, 2024; Palmgren-Neuvonen & Korkeamaeki, 2015; Mengis et al., 2018; Snelson, 2018). Pre-production is the planning stage, production is where the director collects or creates the documents, images, or video for the final product, and post-production includes the assembly of the previously collected media (Palmgren-Neuvonen & Korkeamaeki, 2015; Mengis et al. 2018; Snelson, 2018). Pre-production includes prior knowledge of the director, which may involve evaluating or reflecting on how other media or their own media have been shaped in the past. The production of video contains a set of rules that includes video and sound composition, designing the plan to organize and collect media, and storyboarding the final product (Aaltonen, 2017; Choo et al., 2018; Ketelle, 2018). There are many representational layers to video production, including semiotics, or the making meaning of symbols or the relationship of juxtaposed clips (Lam et al., 2021; Xia, 2024). The role of semiotics plays a role in guiding the production and post-production phases where directors make meaning from the collected materials. The role of semiotics and post-production extends to how the director envisions the final product and how it is then received by the audience, leading to discrepancies in encoding and decoding of semiotics (Bødker, 2016; Conway, 2017; Hall, 2006). Post-production can include distribution of the final product, but this can also be a distinct phase on its own (Bowen, 2024).

Methods and Design:
This study used a descriptive, instrumental single case study approach to investigate the process by which one site implemented video production in its curriculum (Stake, 1995; Yin, 2018). The choice of a case study was made because the particular site was unique in its broad incorporation of video production into the curriculum, and it was prudent to investigate the process by which skills were developed. The bounded case for this study was defined as the process of developing historical skills in the context of the graduate history classroom. Case studies allow for flexibility of data collection to deeply investigate process and flow of a particular site and phenomenon (Yin, 2018). The participants were graduate history students who were 18 or older, who attended a medium to large-sized institution of higher education that offers residential and online course programming. Graduate students were chosen because they were better prepared by the nature of their education to describe historical skills developed through the student-generated video process. The researcher has 10 years in history education and practicing videography skills, creating bias in favor of inclusion of video production in the classrooms. Constructivism guides the framework of the philosophical underpinnings of the study.

Prior to the implementation of the full study, a small pilot of the interview and focus group questions was reviewed to ensure they were appropriate for the study. Participants for the pilot were taken from the population for the general study, as well as a peer review of questions. The pilot resulted in updating questions to be clearly understood, as well as a preliminary understanding of potential themes and patterns that emerged in the full study. For the full study, criterion sampling was used to identify appropriate participants for the study, as well as snowball sampling to reach full saturation. Participants were recruited via email based on whether they had taken a course that contained a video as part of the curriculum. This email was sent out on behalf of the researcher, and potential participants could respond to the request. Roughly 70% of participants were recruited through email, with the rest recruited through snowball sampling from the original responders. Social media was used as a recruitment platform, but no participants were recruited from that source. Participants were asked to sign a consent form, complete a 45-60 minute virtual interview, a 60-minute virtual focus group, submit a guided written response via online form, and a digital artifact in the form of a sample video they produced for their coursework, submitted via online form.

The study contained 10 total participants, 7 of whom participated in the focus group due to scheduling conflicts. Saturation was reached after the 8th and 9th interviews and the second focus group as no new codes were developed from those interviews (Saldaña, 2025). One more interview, journal entry, and video was completed to bring the total participant list to 10. No participant data was excluded. Incentive compensation was offered to all potential participants in the form of a $25 digital gift card. Participants were also given pseudonyms to protect their identities.

The research questions were as follows:
Central Question: How do graduate students describe the development of historical thinking skills through the use of student-generated video in the history classroom?
Sub-question 1: How do graduate students develop the framing skills of historical thinking—historical significance, argumentation, and use of evidence—through student-generated video assessments?
Sub-question 2: How are the primary source evaluation skills—contextualization, corroboration, sourcing, and close reading—developed through the process of graduate students creating their own videos in the history classroom?
Sub-question 3: How do graduate students develop the synthesizing skills of historical thinking—using cause and consequence, continuity and change, and historical perspective-taking—developed through student-generated video in the history classroom?

Results
Data analysis followed the process of case study research and qualitative data coding strategies. Data were analyzed by one researcher, with oversight and audits from the research committee. The researcher was also the key investigator, with review audits coming from colleagues, but no further development of codes and themes from those reviewers. Member checking was completed to verify the themes matched expectations of the participants. The lack of intercoder feedback and discussion was mitigated by design of data collection as well as reflexivity practices found in memos throughout the data collection process (Lincoln & Guba, 1985). Reflexive memos were written during each interview and focus group to guide expectations and surprises from the researcher. These moments helped to separate bias in the data analysis phase.

After data were transcribed from interviews and focus groups, cleaned from written prompts, and videos assessed through an observation rubric, a two-cycle coding strategy was followed. Structural coding served as the first cycle approach, whereby a central framework was applied to the data (Saldaña, 2025). Pattern coding was applied

to further strengthen connections in the data during the second cycle coding process (Saldaña, 2025). All data were checked against memos taken by the researcher to ensure triangulation and validity, as well as a further review from colleagues and member checking of themes. Triangulation was also ensured by the variety of data types, interviews, focus groups, journals, and videos. Data types in different media offered all participants alternative ways to communicate and one way to show evidence of their work through the sample video. A rubric was used to cross-reference all video submissions to allow for an equal position to discuss the products. All data were then presented in the linear-analytic model to connect theory, model of research design, and research results (Yin, 2018).

Participants were divided into levels of comfort and experience with video production, six of which were labeled as beginner, three were labeled as intermediate, and one was labeled as advanced. Data collected through interviews, focus groups, journals, and video analysis all converged to develop the themes. The data was used to corroborate themes, as well as a member check of the findings after the results were outlined. The following four themes and their subthemes were the result of the analysis on historical skills development through student-generated video: framing, evaluation, synthesis, and technical video skills.

Table 1: *Research Participants Experience (N = 10)*

Participant	Degree Track	Content Area	Experience Level with Video
Brittany	Master's	History	Intermediate
Corrine	Ph.D.	History	Intermediate
Fred	Ph.D.	History	Beginner
Jane	Ph.D.	History	Beginner
Jill	Ph.D.	History	Beginner
Liam	Ph.D.	History	Beginner
Megan	Ph.D.	History	Beginner
Michael	Ph.D.	History	Intermediate
Robert	Ph.D.	History	Advanced
Simon	Master's	History	Beginner

Note. Degree track represents highest degree being pursued at time of study. Experience level was assessed through researcher observation during focus groups and self-reports in interviews. Ph.D. = Doctor of Philosophy.

The first theme, framing, described the process by which participants reported organizing and planning their videos. The participants drew on their previous experience with writing history papers and doing research to frame their videos. Fred, one of the participants, reflected, "To me this [in reference to the video assignments] is no different than writing a paper. I must have an introduction/thesis, a body – supporting evidence, and a conclusion." Other participants supported and built on this theme by articulating their perspectives on framing their videos in terms of who their audience was for a particular video, as well as the argument behind it. The role of structure to the video, including the centrality of forming historical arguments was clear, from the participant responses. How each participant collected and used evidence varied from topic to topic, with some just pulling books of the shelf and others digging into primary sources and translated material to understand a very niche topic. They used this information to support their arguments, or even to refine it when it was necessary to adjust it.

Several subthemes emerged to support the development of a fuller picture of framing. Historical significance, argumentation, and use of evidence. Historical significance for some participants was guided by the topic for the class to help them elaborate on a topic for their video. For others, historical significance was an opportunity to link general and particular history to tell a local story while connecting broad, sweeping historical events. Arguments were central to the videos for the participants as they allowed them to focus on a central theme and carry it through each element of the video. Rather than just the structure of content or chronology, they used the argument as the central frame on which they built the rest of their film. The participants generally agreed that they had to make decisions on how they used evidence, what was left in or taken out, in order to produce a clear documentary that met the time constraints.

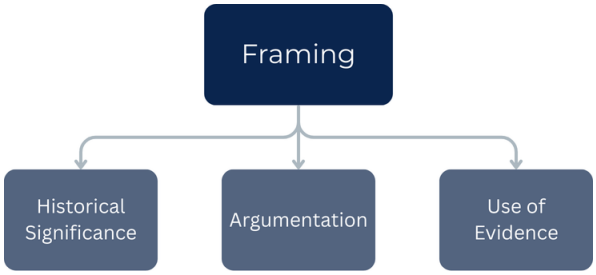


Figure 2.
Note. Image contains a flowchart representing the central theme of framing with its subthemes. Framing consists of the design aspects of a historical film project.

Evaluation was a theme that emerged to describe the process that largely occurred behind the scenes in historical investigation. A lot of the participants described how they would review sources, or even their interviews, within a broader context of the rest of the sources, but that much of the evaluation work rarely makes it on film. From one of the reflective journal entries from Robert, "This type of work required conducting first background research on each individual as well as historical research on their connection we wanted to explore in the film." To paraphrase from Liam, there's a lot more room when writing an essay to elaborate on the differences and similarities between sources, where that may not work as well on film.

The participants covered subthemes of evaluation, including contextualization, corroboration, close-reading, sourcing, and primary sources as visual elements. Most of these align with DIG's (2020) model for analyzing historical thinking skills. Participants discussed the importance of contextualization with any source, as well as corroborating it within the larger pool of information to ensure that the claims could be verified. Sourcing and close reading were important for participants as they allowed them to understand the person behind the source as well as the intent of the author. Some of the considerations of the participants included evaluating the meaning behind words in translations, the location of a site based on conflicting primary records, and how different sources can strengthen arguments by corroborating core claims. Primary sources benefited from becoming tools for inclusion in video, more often than would otherwise be included in most traditional historical research. From Robert's journal, "Film utilizes multiple mediums to engage the viewer. Oral and visual historical content along with sounds and music create an experience that can captivate a wider audience that may not otherwise engage the same material in written form." Using multimodal expressions connected ideas to viewers, but also showed the richness of historical analysis.

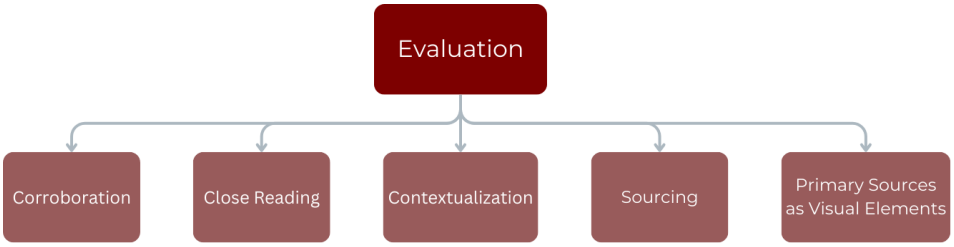


Figure 3.
Note. Image includes a flow chart representing the theme of evaluation with its subthemes. Evaluation represents the analysis phase of historical filmmaking.

Synthesis is a theme that encapsulates the process by which the participants articulated the connections they made in their visual narratives. This was a fascinating theme as it explored the intricacies of not just presenting historical research but connecting and building a narrative to express that research. As Liam wrote in his reflective journal, "The key to organizing video presentations, though, is not the organization itself, but rather how smooth and cogent the transitions are from subtopic to subtopic and point to point." Michael built on establishing a narrative throughline by discussing the way he constructed the narrative of the story through the stages of the battle as it unfolded. He used the phrase "what happened" as a core organizing framework for discussing the events, to then offer an interpretation of those events through historical scholarship.

The following subthemes emerged to explain synthesis: continuity and change, multiperspectivity, cause and consequence, and bias. The subthemes overlapped with the literature from Seixas (2017) and Monte-Sano (2016) to explain the connective tissue of the work of the historian. Continuity and change were a present discussion as

the concept placed historical investigations in a long chain of history, particularly through Jane's comment in the interview

> To make it relevant like show where the path has intersected with the current times or you know why that something that happened back then why you should study it because it led to one of them that led to another, that led to another, that has shaped our world today.

Causality was at the center of historical discussion as it was the catalyst that changed history, and as Robert emphasized, those are the areas that historians gravitate toward when studying the past. Several participants mentioned multiperspectivity as either a collection of different senses or different perspectives from various sources in the past. When varied perspectives were brought in, the participants discussed the issue of bias and ensuring that both historians remain objective or at least discuss the role of bias in their research, while also accounting for it within the texts and documents they reviewed. Corrine summed up bias by stating in the focus group, "you may have your own views on it, but you do have to acknowledge that there are other views."

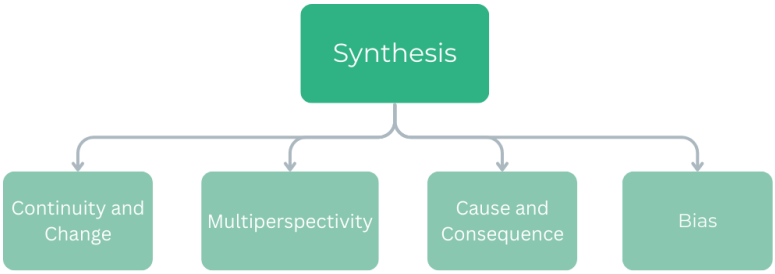


Figure 4.

Note. Image includes a flow chart representing the central theme of synthesis with its subthemes. Synthesis represents the connective elements of drawing a historical film project together.

Beyond the specific historical skills that emerged to align with the video process, there were specific video skills that were recognized by the participants. The theme emerged as a way of highlighting the gap between historical skill and the development of video skills for the students. Many participants learned video skills on their own rather than have a set curriculum to guide them. The participants discussed how the length of the video impacted how they were structuring their ideas as well as how deep they went in the analysis. This was an issue in comparison to writing where much more, according to Jill, "could fit so much more information in a page rather than a timed video presentation. Quality also became a concern for participants as many were novice before making a video for their courses. Some participants focused on self-presentation or the presentation of the talent on screen while others focused on the quality of sources used via green screens. The participants considered how they were communicating their ideas as well as how they were being received. According to Robert,

> It's really just again trying to stick to that outline of how do we want this film, or what do we want this film to achieve? What did we want to portray? How do we want the audience to feel when they're watching this, and what's the message or the takeaway that we want them to have?

His line of questioning showed the tension between the content the director has in their head and how it is received by the audience consuming it.

Fear was also an ever-present factor for some of the participants, as many of them navigated video production for the first time. Some of the fear came from presenting on camera for the first time and growing comfortable with it. They also expressed some stress with filming in-office or on-location. The participants noted that the on-location shots were necessary for historical work, but presented challenges with wind, crowds, weather, and lighting inconsistencies. Robert encapsulated some of the issues with using film equipment: "Moreover, because shoots cannot often be repeated, all the necessary film materials must be shot the day of the interview; one cannot go back to add questions, follow-ups, of B-roll material that would enhance the quality of the film." The stress with working with filmmaking equipment extended to using video editing programs. Jane encapsulated some of the struggle with the equipment by stating that: "editing is tedious and challenging but an essential part of every project." The consensus among participants was that there are many skills outside of the historical realm that they had to embrace to produce a video that communicated their research.

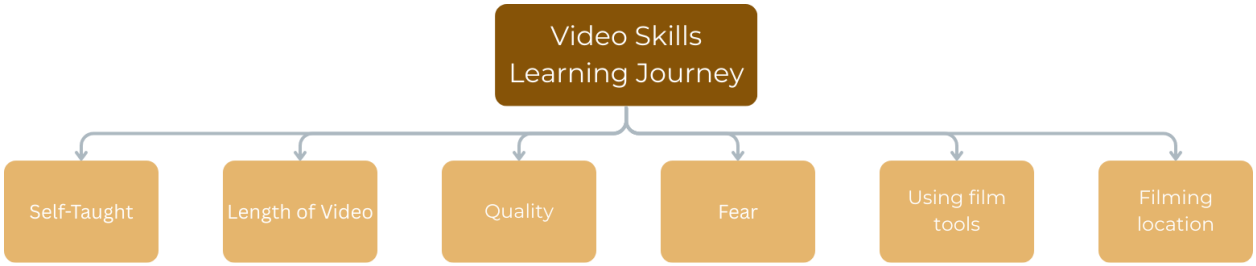


Figure 5.

Note. Image includes a flow chart representing the central theme with subthemes. The theme addresses video skills and challenges from the perspective of students.

While most of the participants noted positive experiences with video, despite frustration, one participant, Jill, was more negative regarding the process. Most of these comments were centered around the structure of the classroom and perceptions that the assignments were easier to grade, which was not related to the topic of the rest of the study. This negative perception, in turn, spilled into frustration with the use of video as an assessment tool. Understanding this negative perspective, coupled with some fear and trepidation about the use of film in the classroom, was useful as a foil to the rest of the themes and data collected.

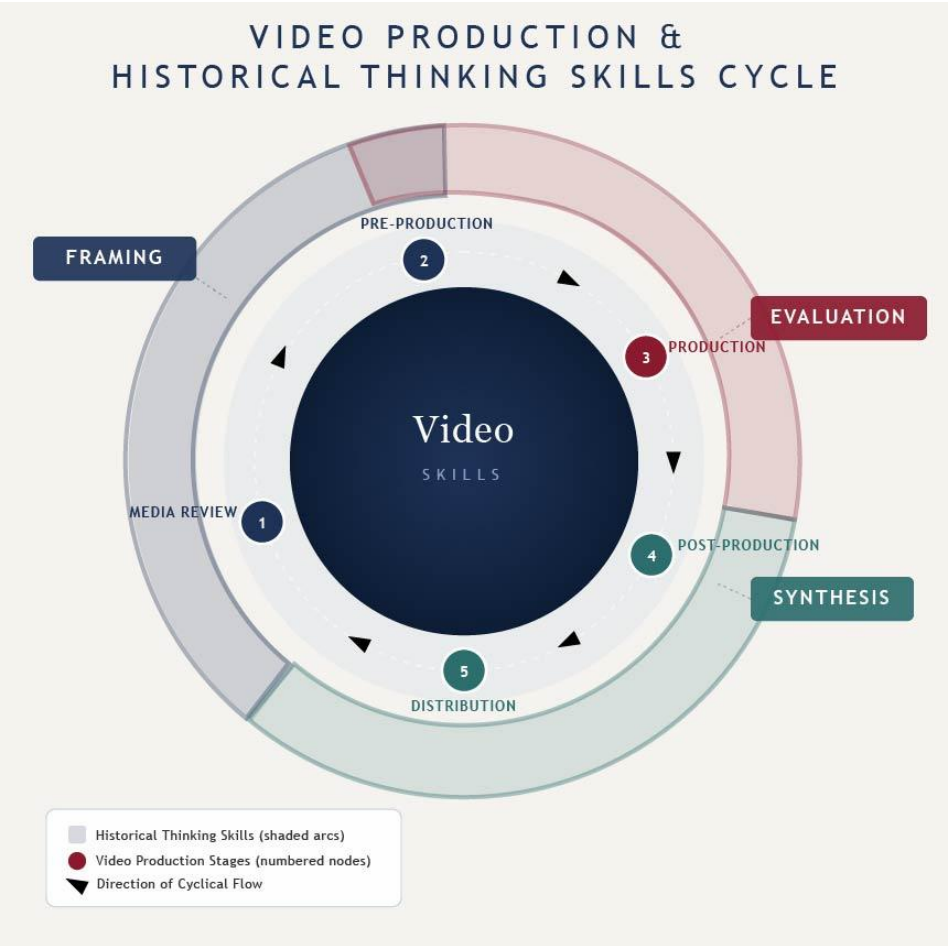


Figure 6.

Note. Image developed to visualize the position of historical skills relative to video skills within the video production process. Developed by author.

Discussion

This study brought together historical thinking and video production to align the cognitive skills of analysis with the technical skills of producing films. The results showed considerable overlap in the development of videos in the classroom, with advanced history students articulating historical skills through the process. It is useful at this point to draw attention back to the video production process, starting with media review, going through pre-production, production, post-production, and distribution (Figure 6). The core themes of this study serve as a

blanketing mechanism to encapsulate these processes and highlight areas of overlap between the work of historical analysis within video production.

Thematically, framing as the first of the broad topics connects across disciplines to establish a more robust framework for understanding and utilizing film as a medium for inquiry (Aaltonen, 2017; Arruabarrena, et al., 2019). Framing includes elements of media literacy to foster a full understanding of the type of media being produced, from documentaries to dramatic works, thereby guiding discussion toward purpose, audience, and argument (Monte-Sano, 2016). By threading the line between media literacy, historical analysis, and pre-production, framing encompasses a broad swath of the planning process of filmmaking (Madariaga et al., 2021; Mengis et al., 2018; Monte-Sano, 2016; Seixas, 2017; Snelson, 2018). The design-centric focus of framing allows for more reflection and inquiry before more work begins. The findings of the study agree with and link disparate aspects of the literature to draw a connection between the historical process of framing research to filmmaking skills involved with pre-production (Aaltonen, 2017; Arruabarrena, et al., 2019; Seixas, 2017; Xia, 2024).

As students continue the cycle of development toward production, they encounter an overlap between framing and evaluation. Where framing serves as a place to plan, ideate, and organize, evaluation serves as a place to parse the data that has been collected and to evaluate it for its purpose within the story (Caulfield & Wineburg, 2023; Jones-Jang et al., 2021; Park et al., 2021; Yuan et al., 2021). Students in this phase must evaluate the sources that they have either for authenticity, historical provenance, weight to support claims, or to see how they will fit in the narrative (DIG, 2020). The findings of the study place emphasis on the connections between the evaluative historical skills and the process of production, where directors collect and organize their shots, visuals, and other materials (Madariaga et al., 2021). Participants identify skills in the evaluative phase as either comfortable for most of the historical skills, or confusing if they were on the technical side of video production (Snelson, 2018). This expression from participants aligns with other studies that focus on student experience and some of the reservations they have with using film equipment for the first time (Arruabarrena, et al., 2019; Madariaga et al., 2021).

Rounding the last third of the video production cycle features students engaging in post-production and distributing their films. For historical films, the post-production cycle features connecting sources together to create the argument that was planned in the pre-production stage. Rather than merely design the narrative, synthesis of the materials is where the story takes shape. To do this student assign meaning to texts, clarify it with images, and highlight important arguments with audio clips (Conway, 2017; Hall, 2006; Monte-Sano, 2016; Seixas, 2017). The study builds on video production by adding a layer of historical analysis not readily apparent in the literature (Seixas, 2017). Meaning-making exists as a topic within the broader scope of multimodal design, but this study augments the skill to layer in historical thinking (Aaltonen, 2017; Bowen, 2024; Madariaga et al., 2021). As participants articulated, editing and post-production can pose technical challenges, which aligns with the discussion of literature where students express little prior knowledge of the video editing software or process (Madariaga et al., 2021; Palmgren-Neuvonen & Korkeamäki, 2015).

The findings of the study also add to the literature that filmmaking is an iterative process that continues as students move on. As the spiral curriculum of Bruner (1977) indicates, the more a student engages in a skill or approaches a content area, the more they learn by seeing a different aspect of the subject. Students in video production will finish a project, reflect on its outcome, and start again with another project through this reflection to strengthen video skills (Conway, 2017; Engeström, 2001; Hall, 2006). Similarly with history, seeing the topic from unique angles and through multiple sources allows students to develop more nuanced takes on content-level history (Monte-Sano, 2016; Seixas, 2017). This cyclical nature of filmmaking is not new, nor is the spiral curriculum, but the connection of the domains is not predominantly portrayed in the literature (Snelson, 2018). From a theoretical perspective, the results draw on Cognitive Constructivism to provide a mental framework and on Activity Theory to provide clarity on the importance of the tools, rules, and community in which students develop their final video products. The internal processes, represented by historical thinking skills, and the external processes, represented by the tools and rules of filmmaking, join theory, literature, and findings into one cogent framework for thinking of future applications of student-directed videos in the classroom.

The implications of the study build a framework for policy considerations as well as practice changes or adoption. First, it builds the literature further to connect disparate fields of historical thinking and video production. Connecting the literature to evaluate the design components of video production assessments to build historical skills sets up further study and implementation across various age ranges. For policy consideration, school districts can recommend different modalities, including video production, to teachers. They can also use it to leverage additional funding for equipment for the classrooms. As national organizations evolve, they can be used to leverage recommendations on best practices or policy surrounding safe filmmaking practices with respect to

privacy and protection for students in the K12 environment. Currently, ISTE Standards support the teacher and student in the development of videos as they give students hands-on experience with different modalities for communication, as well as giving them valuable technology training (ISTE, 2024a; ISTE, 2024b).

For practice consideration, there would be additional professional development needed as video production includes some elements of learning the technology to learn with or be a step ahead of the students as they are completing their assignments. Structuring learning for the educators as well as the students would be the first step in providing updates to practice. Teachers would also need to consider their resources before committing to adding to the curriculum. If there are limited resources, then the practical inclusion of video in the classroom may take the form of group work and design rather than full implementation of video. This involves a smaller-scale implementation of the work, without the high cost of technology such as cameras and non-linear editing programs. Additionally, more knowledge about cheap or free resources would be necessary for teachers, as well as a rudimentary knowledge of fair use and copyright laws for the use of primary sources and videos.

This study has lasting implications for educational technology, including promotion of multimodal design within the historical field, where writing has been the sole focus of communication for centuries (Lukacs, 2011). Learning the grammar of filmmaking allows students to broaden their perspectives and deepen their understanding through a multilayered understanding of a topic (Bowen, 2024; Bruner, 1961; Caulfield & Wineburg, 2023; Jones-Jang et al., 2021). Thinking visually, conceptually, and argumentatively provides a complex product that is produced through deep thinking. History as a discipline is often slower to adopt technology, but the use of film allows history students to learn the skills of documentarians in the field to shift the nature of their inquiry beyond what can often be limited to text-based sources (Barnouw, 1993; Bowen, 2024; Lukacs, 2011). Finally, digital and AI literacy are enhanced through engaging with various media to develop interpretations of history (Caulfield & Weinburg, 2023; Monte-Sano, 2016; Xia, 2024). Requiring students to be critical of images, either real or generated, provides an opportunity for students to practice evaluative skills that in turn fuel other skills of development of argument and synthesizing the evidence into clear interpretations (Madariaga et al., 2021; Seixas, 2017; Snelson, 2018; Wu & Zhang, 2025).

Limitations and Delimitations:

This study contained limitations first in timing of the study and then in impact of programmatic change. The study took place during the COVID-19 pandemic when many residential programs were forced online. This impacted some cognates and programs which included on-site video production and travel for history class credit. Because of the pandemic, these programs were shuttered, not to resume after residential courses were offered again. This impacted recruitment, caused delays, and altered the original design of the study to fit a more flexible participant pool.

The study itself was intentionally delimited to include only graduate students of history to ensure that they were able to clearly articulate historical skills within a unique paradigm of video production. Undergraduate programs did not include the breadth of video production expected for students to adopt higher-level reasoning as well as deeper research and scholarship. The definitions of the historical skills and the availability of the population were the largest reason to delimit to just graduate students in history programs. Those who were film majors alone were not considered as they did not have the background of historical skills required to complete the study.

From this research, there are opportunities to extend similar case studies or qualitative studies to the K12 and undergraduate history classrooms. Future studies would focus on historical thinking skills and video production, but change the context to undergraduate, secondary, and primary education. The goal is to understand the perspectives that these students can articulate regarding the varied applications of skills, as well as to see the efficacy of the assessment design on their skill development and critical thinking. Broadly, it is crucial to establish student-generated video within the Historical Assessments of Thinking (Smith et al., 2019). Using the Historical Assessments of Thinking instrument will allow for generalizability about areas that are more or less prevalent.

Conclusion

With shifts in technology making it possible for students to become directors of their own films and with the medium being pervasive for broad communication, it is imperative to position filmmaking within more curricula. History is a discipline that is often reluctant to change its methodology or presentation, yet it is also a field where works of all types – books, podcasts, and videos – are widely accessible to the public beyond the classroom. Engaging with history in ways that access its storytelling aspects and combining that with the medium of filmmaking gives teachers an opportunity to build content-level skills, technical competence, and create artifacts that others in the community can enjoy. This level of hands-on experience is good for students to immerse themselves in the world they are studying, and it has the additional impact of keeping human-centric design when

many other fields face persistent threats from unconstrained AI. History provides the skills necessary to combat some of the ills of AI, while also acknowledging its potential for good. This case study analyzed the process of student-directed filmmaking to build a useful conceptual model based on the theories of Cognitive Constructivism and Activity Theory to structure historical thinking through the rules and tools of videography. Filmmaking in the history classroom thus provides the nexus of learning where students can be challenged pedagogically, rewarded creatively, and leave with a new set of interpretative and communication skills upon completion of a course of study.

Acknowledgements:
This document was constructed as part of the findings of a dissertation study. Dissertation supervisors are acknowledged for their time and guidance of the study. In the final stages of production of this particular article, Artificial Intelligence models were used to provide triangulated feedback to strengthen the writing, but not to generate text for the article. The AI models were used as a preliminary peer-review before submission. The image of themes was refined by AI from a sketch from the author to produce a more legible design without access to more nuanced graphic design software. This chain of development can be provided upon request. There are no known conflicts of interest with the researcher.

References

Aaltonen, J. (2017). Script as a hypothesis: Scriptwriting for documentary film. *Journal of Screenwriting*, *8*(1), 55-65. https://doi.org/10.1386/josc.8.1.55-1

Arruabarrena, R., Sánchez, A., Blanco, J. M., Vadillo, J. A., & Usandizaga, I. (2019). Integration of good practices of active methodologies with the reuse of student-generated content. *International Journal of Educational Technology in Higher Education*, *16*(1). https://doi.org/10.1186/s41239-019-0140-7

Arruabarrena, R., Sánchez, A., Domínguez, C., & Jaime, A. (2021). A novel taxonomy of student-generated video styles. *International Journal of Educational Technology in Higher Education*, *18*(1), 68. https://doi.org/10.1186/s41239-021-00295-6

Barnouw, E. (1993). *Documentary: A history of the non-fiction film*. Oxford University Press.

Bowen, C. (2024). *Grammar of the edit.* Routledge.

Bødker, H. (2016). Stuart Hall's encoding/decoding model and the circulation of journalism in the digital landscape. *Critical Studies in Media Communication*, *33*(5), 409-423.

Breakstone, J., McGrew, S., Smith, M., Ortega, T., & Wineburg, S. (2018, March). Why we need a new approach to teaching digital literacy: To assess the credibility of the information they find online, students shouldn't start with a close reading of the given website. Rather, they should turn to the power of the web to determine its trustworthiness. *Phi Delta Kappan*, *99*(6), 27. https://doi.org/10.1177/0031721718762419

Bruner, J. S. (1966). *Toward a theory of instruction*. Harvard University Press.

Bruner, J. S. (1977). *The process of education.* Harvard University Press.

Bruner, J. S. (2006). *In search of pedagogy, volume I: The selected writings of Jerome Bruner, 1957-1978*. Routledge.

Carr, E. H. (1961). *What is history?* Random House, Inc.

Caulfield, M., & Wineburg, S. (2023). *Verified: How to think straight, get duped less, and make better decisions about what to believe online*. University of Chicago Press.

Choo, Y. B., Abdullah, T., & Nawi, A. M. (2018). Digital storytelling to unlock reflective practice in the classroom. *Asian Social Science*, *14*(11), 52. https://doi.org/10.5539/ass.v14n11p52

Conway, K. (2017). Communication is translation, or, how to mind the gap. *Palabra Clave*, *20*(3), 622-644. https://doi.org/10.5294/pacla.2017.20.3.2

Dewey, J. (1910). *How we think.* D.C. Heath & Co Publishers.

Digital Inquiry Group [DIG]. (2020). Historical thinking chart. *Stanford History Education Group.* https://www.inquirygroup.org/history-lessons/historical-thinking-chart

Engeström, Y. (1987). *Learning by expanding: An activity-theoretical approach to developmental research*. Orienta-Konsultit Oy.

Engeström, Y. (2001). Expansive learning at work: Toward an activity-theoretical reconceptualization. *Journal of Education and Work*, *14*(1), 133-156.

Fea, J. (2013). *Why study history?: Reflecting on the importance of the past*. Baker Academic.

Gaddis, J. L. (2004). *Landscape of history: How historians map the past*. Oxford University Press.

Gagné, R. M. (1970). *The conditions of learning*. Holt, Rinehart & Winston.

Guzmán-Simón, F., García-Jiménez, E., & López-Cobo, I. (2017). Undergraduate students' perspectives on digital competence and academic literacy in a Spanish University. *Computers in Human Behavior*, *74*, 196-204. https://doi.org/10.1016/j.chb.2017.04.040

Hall, S. (2006). Encoding/Decoding. In M. G. Durham & D. M. Kellner (Eds), *Media and cultural studies* (pp. 164-173). Blackwell Publishing.
Huang, Q., & Xia, S. (2024). Preparing learners for digitally mediated academic communication: Digital multimodal practice in students' knowledge dissemination videos. *Journal of English for Academic Purposes*, *71*, 101429. https://doi.org/10.1016/j.jeap.2024.101429
Husbye, N. E., & Vander Zanden, S. (2015). Composing film: Multimodality and production in elementary classrooms. *Theory Into Practice*, *54*(2), 109–116. https://doi.org/10.1080/00405841.2015.1010840
International Society for Technology Education [ISTE]. (2024) ISTE standards: For educators. https://iste.org/standards/educators
International Society for Technology Education [ISTE]. (2024). ISTE standards: For students. https://iste.org/standards/students
Jones-Jang, S. M., Mortensen, T., & Liu, J. (2021). Does media literacy help identification of fake news? Information literacy helps, but other literacies don't. *American Behavioral Scientist*, *65*(2), 371–388. https://doi.org/10.1177/0002764219869406
Karasavvvidis, I. (2019). Students' use of digital video effects in multimodal compositions: An exploratory study. *Journal of Visual Literacy*, *38*(3), 221-244. https://doi.org/10.1080/1051144x.2019.1611697
Ketelle, D. (2018). What is storytelling in the higher education classroom? *Storytelling, Self, Society*, *13*(2), 143-150. https://doi.org/10.13110/storselfsoci.13.2.0143
Lam, W. S. E., Smirnov, N., Chang, A. A., Easterday, M. W., Rosario-Ramos, E. M., & Doppelt, J. C. (2021). Multimodal voicing and scale-making in a youth-produced video documentary on immigration. *Research in the Teaching of English*, *55*(4), 340–368. https://doi.org/10.58680/rte202131256
Lowenthal, D. (1987). *The past is a foreign country.* Cambridge University Press.
Lukacs, J. (2011). *The future of history*. Yale University Press.
Madariaga, L., Nussbaum, M., Gutiérrez, I., Barahona, C., & Meneses, A. (2021). Assessment of user experience in video-based learning environments: From design guidelines to final product. *Computers & Education*, *167*, https://doi.org/10.1016/j.compedu.2021.104176
Masanet, M. J., Guerrero-Pico, M., & Establés, M. J. (2019). From digital native to digital apprentice. A case study of the transmedia skills and informal learning strategies of adolescents in Spain. *Learning, Media and Technology*, *44*(4), 400-413. https://doi.org/10.1080/17439884.2019.1641513
McGahan, W., Ernst, H., & Dyson, L. E. (2016). Individual learning strategies and choice in student-generated multimedia. *International Journal of Mobile and Blended Learning*, *8*(3), 1-18. https://doi.org/10.4018/IJMBL.2016070101
Mengis, J., Nicolini, D., & Gorli, M. (2018). The video production of space: How different recording practices matter. *Organizational Research Methods*, *21*(2), 288-315. https://doi.org/10.1177/1094428116669819
Miller, S. M. (2013). A research metasynthesis on digital video composing in classrooms: An evidence-based framework toward a pedagogy for embodied learning. *Journal of Literacy Research*, *45*(4), 386-430. https://doi.org/10.1177/1086296X13504867
Monte-Sano, C. (2016). Argumentation in history classrooms: A key path to understanding the discipline and preparing citizens. *Theory into Practice*, *55*(4), 311-319. https://doi.org/10.1080/00405841.2016.1208068
Nichols, T. P., & Stornaiuolo, A. (2019). Assembling ″digital literacies": Contingent pasts, possible futures. *Media and Communication*, *7*(2), 14-24. https://doi.org/10.17645/mac.v7i2.1946
Orak, C., & Turan, Z. (2024). Using artificial intelligence in digital video production: A systematic review study. *Journal of Educational Technology and Online Learning*, *7*(3), 286–307. https://doi.org/10.31681/jetol.1459434
Palmgren-Neuvonen, L., & Korkeamäki, R. L. (2015). Teacher as an orchestrator of collaborative planning in learner-generated video production. *Learning, Culture and Social Interaction*, *7*, 1-11. https://doi.org/10.1016/j.lcsi.2015.09.002
Park, H., Kim, H. S., & Park, H. W. (2021). A scientometric study of digital literacy, ICT literacy, information literacy, and media literacy. *Journal of Data and Information Science*, *6*(2), 116–138. https://doi.org/10.2478/jdis-2021-0001
Piaget, J. (1977). *The essential Piaget* (H. E. Gruber & J. J. Vonèche, Eds.). Basic Books, Inc.
Rainville, K. N., & Gordh, B. (2016). Toward a narrative classroom: Storytelling, media, and literacy. *YC Young Children*, *71*(4), 76-81.
Reyna, J., & Meier, P. (2018). Learner-generated digital media (LGDM) as an assessment tool in tertiary science education: A review of literature. *IAFOR Journal of Education*, *6*(3), 93-109. https://doi.org/10.22492/ije.6.3.06
Saldaña, J. (2025). *The coding manual for qualitative researchers*. SAGE Publishing Inc.
Scolari, C. A. (2019). Beyond the myth of the "digital native." *Nordic Journal of Digital Literacy*, *14*(03-04), 164-174. https://doi.org/10.18261/issn.1891-943x-2019-03-04-06

Seixas, P. (2017). A model of historical thinking. *Educational Philosophy and Theory*, *49*(6), 593-605. https://doi.org/10.1080/00131857.2015.1101363

Smith, M., Breakstone, J., and Wineburg, S., (2019). History assessments of thinking: A validity study. *Cognition and Instruction*. doi: 10.1080/07370008.2018.1499646

Snelson, C. (2018). Video production in content-area pedagogy: A scoping study of the research literature. *Learning, Media and Technology*, *43*(3), 294-306. https://doi.org/10.1080/17439884.2018.1504788

Stake, R. E. (1995). *The art of case study research*. Sage Publications.

Stapleton, L., & Stefaniak, J. (2019). Cognitive constructivism: Revisiting Jerome Bruner's influence on instructional design practices. *TechTrends*, *63*(1), 4-5. https://doi.org/10.1007/s11528-018-0356-8

Tiernan, P., & Farren, M. (2017). Digital literacy and online video: Undergraduate students' use of online video for coursework. *Education and Information Technologies*, *22*(6), 3167-3185. https://doi.org/10.1007/s10639-017-9575-4

Xia, S. (2024). Incorporating digital multimodal composition in content teaching: A multimodal analysis of students' legal popularization videos. *Journal of Second Language Writing*, *66*, 101163. https://doi.org/10.1016/j.jslw.2024.101163

Wu, D., & Zhang, J. (2025). Generative artificial intelligence in secondary education: Applications and effects on students' innovation skills and digital literacy. *PLOS One*, *20*(5), e0323349. https://doi.org/10.1371/journal.pone.0323349

Yin, R. K. (2018). *Case study research and applications: Design and methods*. SAGE Publications, Inc.

Yuan, Y.-H., Liu, C.-H., & Kuang, S.-S. (2021). An Innovative and Interactive Teaching Model for Cultivating Talent's Digital Literacy in Decision Making, Sustainability, and Computational Thinking. *Sustainability*, *13*(9), 5117. https://doi.org/10.3390/su13095117

APPENDIX A: INTERVIEW QUESTIONS

1. Good morning/afternoon, I'm glad you could join me for this interview. I want to start off getting to know your background with history. How did you decide to study history at the graduate level?
2. I would like for you to discuss the video that you made for your history class. Can you give a brief description of your video?
3. What is your own experience with making videos?
4. What do you see as a difference between doing a video and writing a paper?
5. The next few questions will cover a similar topic from different stages. Can you explain how you planned and researched your videos?
6. Can you describe the process of filming your video?
7. Can you explain how you assembled and finalized your video?
8. What are some skills that you developed through this assignment?
9. In the next few questions, we are going to look at historical skills specifically. What are some skills that a historian must have to interpret the past?
10. How did you incorporate and analyze primary sources in your video?
11. How did you show historical cause and effect through your video?
12. How did you identify your topic as historically significant?
13. Describe how you formed your argument or interpretation in your video.
14. Through your video, how did you incorporate multiple historical perspectives to interpret the past?
15. Can you describe how you contextualized the sources used in your video?
16. What guided your research as you created the videos?
17. What do you think would improve the program of making videos in the classroom?
18. As we are wrapping up, what else would you like to add about your experience with student-directed video?

APPENDIX B: FOCUS GROUP QUESTIONS

1. Can you describe the video project that you completed for your history class?
2. How did you approach organizing and planning research?
3. How did making a video impact the way you research in your history class?
4. What are some things that inhibited creating a video for your history class?
5. How did producing a video compare to writing a paper, specifically as far as your research process is concerned?
6. What historical skills are you aware of now that you have completed a video assignment?
7. To wrap up our discussion here, what topics would you like to cover that we have not already addressed?

APPENDIX C: REFLECTIVE JOURNAL QUESTIONS

Instructions: For the following prompts, please provide a reflection that is between 500 – 1,000 words for each question when you are in each phase of your video specified in the prompt. For example, when you are preparing and planning your video write your reflection on prompt 1. Retain all journals until you have completed all three. You will submit them to a form to collect your work.

1. Please discuss your thoughts as you have searched for research in preparation for this project. Be candid about any stumbling blocks or differences you may experience from typical history assignments.
2. Please discuss your thoughts on the process of making a video. What did you discover was an issue for you that you did not expect? How do you see this process in the scope of learning in the field of history?
3. Please discuss your thoughts on the process of editing and structuring the story of your video. What was your experience with thinking of how to organize the material you collected? How does it compare to typical history assessments?

APPENDIX D: PILOT STUDY

Procedures

Faculty and potential participants for the study were contacted to complete a review of questions used in the study. Both a faculty and expert review as well as a potential participant review were necessary to see the efficacy of the questions. Departmental approval was ascertained before the pilot study commenced to ensure that the correct students and faculty were contacted with permission. As this pilot study was indicated through the IRB as a review of questions, the procedures for some question asking was maintained, but the questions were workshopped to improve clarity. Experts and participants were asked to participate in interviews, focus groups, and to complete reflective journals as they are completing video assignments. This information was used to understand what information the reader saw in each question. This information, as well as a review of the questions with each participant after was used to clarify any misunderstandings in the process of posing questions.

From the pilot study a few issues were raised with the questions in the interview and focus group. First with the interview, several of the questions that focused on skill development, namely questions 9-14 seemed to be confusing for participants. These questions were adjusted heavily to specify the specific skill to clarify in the context of video. Additionally, some answers were broadly on the subject of history instead of specifically on the video that was created and how historical skills were used for this, so question 2 was adjusted to frame the rest of the questions in the context of that student-generated video. One question was removed entirely as it seemed that it did not fit and only confused participants, this was question 16. Other than these mentioned, the other questions were only minimally changed for clarity or otherwise not changed.

Regarding the focus group questions, these were clearer and the dialog created amongst participants created good conversation. There were not too many needing corrections for clarity, but there were a few questions that did cause students to stumble. As it was difficult for students to refer back to before the class to know what skills they had, questions 6 and 7 were combined to focus solely on the skills that students indicated that they developed through the process of making their video. As with the interview questions, the focus did not seem to be on a specific video during the pilot study, so the first question of the focus group was adjusted to allow for the participants to discuss the content of their student-directed video. This would allow for a clearer framing on the topic of the video and the participants would be able to refer to the video to clarify historical skills used in the production of the video.

Furthermore, the background of the participation seemed to dictate how they answered the question. Since the study focuses on graduate students, there is a disparity between experiences of the individuals. Some had far more video experience, teaching expertise, or content knowledge than others and because of that experience leaned on that to answer some questions instead of focusing on the subject matter expected from the question. This information was useful heading into the research study to understand how to approach different backgrounds and what to expect from each. For example, those who had more experience with film equipment focused their attention on the skills of video instead of the skills of historical thinking. As this study considers video production a tertiary aspect of the study, these answers were not as useful as ones that clarified the historical thinking process in video.

Interview Questions

1. Good morning/afternoon, I'm glad you could join me for this interview. I want to start off getting to know your background with history. How did you decide to study history at the graduate level?
2. I would like for you to discuss your journey in making a video as a part of the assignments in this course.
3. What is your own experience with making videos?
4. What do you see as a difference between doing a video and writing a paper?
5. The next few questions will cover a similar topic from different stages. What structures do you identify that are in place to guide you through the process of planning videos?
6. What arrangements do you see in place to guide you through the process of filming videos?
7. What structures do you see in place to guide you through the process of editing videos?
8. What are some skills that you hope to build or have already developed through this assignment?
9. In the next few questions, we are going to look at historical skills specifically. What are some skills that a historian must have to interpret the past?
10. In the process of making this video, how did you approach primary sources?
11. How would you describe the process of approaching cause and effect in making this video?
12. How would you address issues of attributing historical significance in forming your argument and interpretation in making a video?
13. How do you see your video helping in developing skills of multiperspectivity, or seeing multiple perspectives of the past?

14. How do you see fitting your video in context, both in time and place, in making your video?
15. What guided your research as you created the videos?
16. Regarding your expectations, what do you see as the expectation of you in developing your own videos?
17. What do you think would improve the program of making videos in the classroom?
18. As we are wrapping up, what else would you like to add about your experience with student-directed video?

Focus Group Questions

1. Overall, in this process, what did you learn through the process of making your video?
2. How did you approach organizing and planning research?
3. How did this approach impact the way you research in your history class?
4. What are some things that you were inhibited by while producing videos for history classes?
5. How does producing a video compare to writing a paper, specifically as far as your research process is concerned?
6. I'd like to see what each of you brought into this project as far as skills go. What are some of the historical skills that you were aware of before this project?
7. What skills are you aware of now?
8. To wrap up our discussion here, what topics would you like to cover that we have not already addressed?

Reflective Journal Questions

Instructions: For the following prompts, please provide a reflection that is between 500 – 1,000 words for each question when you are in each phase of your video specified in the prompt. For example, when you are preparing and planning your video write your reflection on prompt 1. Retain all journals until you have completed all three. You will submit them to a form to collect your work.

1. Please discuss your thoughts as you have searched for research in preparation for this project. Be candid about any stumbling blocks or differences you may experience from typical history assignments.
2. Please discuss your thoughts on the process of making a video. What did you discover was an issue for you that you did not expect? How do you see this process in the scope of learning in the field of history?
3. Please discuss your thoughts on the process of editing and structuring the story of your video. What was your experience with thinking of how to organize the material you collected? How does it compare to typical history assessments?

APPENDIX E: VIDEO ARTIFACT DATA ANALYSIS RUBRIC

Element of Historical Thinking	**Evident** (3 Points)	**Developing** (2 Points)	**Missing** (1 Point)
FRAMING			
Historical Significance (Seixas, 2017).	Carefully frames the argument of the video around a central theme. Expresses the theme in the light of general and particular history. General history – Broad aspects and themes. Particular is the specific event in focus.	Has a main idea but does not show it clearly or misses part of showing themes in light of general and particular history. May contain broad discussion and not a specific discussion, or vice versa.	Has a topic within the video, but the argument is not clearly defined. Cannot understand the relationship between the broad history and the narrow history as presented.
Use of Evidence (Monte-Sano, 2016; Seixas, 2017).	Uses a blend of primary and secondary sources to underpin and support argument. Understand the limitations of each source used.	Relies more on secondary sources and less on primary sources for the presentation. Understands some of the limitations of sources, but not all.	Relies on secondary sources, textbook, or encyclopedia articles. Does not consult primary texts or present understanding of the limits of sources.
Argumentation (Monte-Sano, 2016).	Lines out argument and the points borne out through the evidence collected. Builds strong argument through use of evidence. Asks questions that can be answered fully.	Has a good argument, but may not have enough evidence to support it. Builds an argument, but evidence either does not support or is missing some parts of the question to be answered.	Has a poorly structured argument and does not supply much evidence to support it. Does not take a critical view and presents argument as a matter of fact.
EVALUATING			
Corroboration (Wineburg, 2001; SHEG, 2020).	Evaluates source within the confines of other sources in the field. Chooses a broad array of sources to show other perspectives and evidence.	Evaluates source through only a few sources. Gives some other examples of sources, but not in-depth.	Takes the source at face value without looking for other evidence.
Close Reading (Wineburg, 2001; SHEG, 2020).	Reads into the intent that the author puts into the text. Takes seriously the meaning behind the words on paper, print, or digital release.	Does a surface level read of the intent of the author. Addresses the idea of analyzing deeper but does not evaluate much deeper meaning or intent.	Reads a source at face value. Does not go into further analysis to think about the intent of the author.
Contextualization (Wineburg, 2001; SHEG, 2020).	Evaluates sources in the context of the time in which they are written. Evaluates this as a background to understand intent and environment of author	Addresses some of the context or addresses the wrong context of the document. Develops part of the picture of the time in which the document is written.	Does not add a meaningful backdrop to the context of documents.
Sourcing (Wineburg, 2001; SHEG, 2020).	Establishes knowledge about the author to clarify intent, ambitions and trustworthiness of author.	Addresses tertiary issues of sourcing, including websites sources, trustworthiness of archives, but not the source itself.	Does not address the trustworthiness of the source. Accepts the description from the author with no other thought of further analysis.

SYNTHESIZING			
Cause and Consequence (Seixas, 2017).	Weaves narrative structures to explain what happened in the past. Show that actions have other actions precipitating from them and the connection in human decision-making.	Somewhat addresses the actions of the past and displaying what happened. Sometimes approaches it from a standpoint of this happened, then this happened, without expressing why or showing the process of decision-making.	No clear thread of action displayed. Difficult to see the human agency involved with decisions of the past.
Continuity and Change (Seixas, 2017).	Weaves narrative to show shows that events do not neatly fall within periods that are defined as many things are occurring at once. Clearly exhibits that history is not a monolith of constant change and progress, but one that involves choices and often involves discontinuity	Weaves a narrative of continual progress. Somewhat addresses issues of continuity and choices but does not fully express or show this. Somewhat addresses issues of multiple events occurring simultaneously.	Does not address issues of continuity. Either sees decisions without context or sees history as one fluid sweep with no interruptions.
Multiperspectivity (Seixas, 2017).	Connects understanding by using multiple perspectives in the past to include more than several points of view in an analysis. Takes into consideration several perspectives as narrative is constructed.	Addresses the existence of other perspectives but does not include them completely in the interpretation. Does not address other points of view in a way that strengthens the narrative.	Does not address multiple perspectives to build an interpretation. Focuses only on one side of the story.

Scopus Database-Based Bibliometric Analysis: Academic Studies/Research on Twice Exceptionality and Giftedness

Sabit MENTEŞE (Prof. Dr.)
Munzur Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Siyaset Bilimi ve Kamu Yönetimi Bölümü, TUNCELİ, TÜRKİYE
ORCID ID: https://orcid.org/0000-0003-4901-4481,
Mail: smentese@munzur.edu.tr, GSM: (0)539 346 86 38

Sema BEKLER (Science Expert)
Munzur Üniversitesi Lisansüstü Eğitim Enstitüsü Öğrenci
ORCID ID: https://orcid.org/0009-0002-2152-2767,
Mail: ksy.semabekler@gmail.com, GSM: (0)5056360906

ABSTRACT
In recent years, the concept of "Twice Exceptional" has been one of the most important research topics in educational psychology and special education. Twice Exceptional generally refers to individuals who possess high cognitive abilities, creative thinking skills, etc., and who also exhibit specific learning difficulties alongside attention deficit disorder, autism spectrum disorder, social and emotional development disorder, speech disorder, attention disorder, social adjustment disorder, etc. The examination of various aspects of studies conducted on the subject in the literature is expected to contribute to educational approaches targeting two different groups of individuals, enabling educators, families, and therapists to respond more effectively to the needs of these students. In this context, the study was designed as a case study in order to examine the sample obtained from the Scopus database in terms of bibliometric parameters and determine the current situation. In this study, which employs a bibliometric analysis approach, data were obtained from the Scopus database covering the period 1995-2024. The outputs obtained from the database were mapped using VOSviewer (version 1.6.20), a modern program that enables the accurate and understandable interpretation of large datasets. A search conducted in the Scopus database on January 11, 2025, using the keywords "twice-exceptional" OR '2e' OR "dual exceptional" OR "multiple exceptional" AND "gifted" yielded a total of 354 documents, including 236 articles, within the search period of 1995-2024. According to the findings of the study, the year with the highest number of studies on giftedness and excellence indexed in the Scopus database was 2015, with 40 articles. Of the 236 articles comprising the study sample, 92.80% were written in English. When examining the most cited articles in the study, Foley Nicpon, M. (2011) ranked first with 188 citations, Reis, S. M. (2014) ranked second with 155 citations, and Assouline, S. G. (2010) ranked third with 107 citations. Within the scope of the study, Foley-Nicpon, Megan, with 19 documents and 723 citations, Assouline, Susan G., with 14 documents and 553 citations, and Baum, Susan M., with 4 documents and 205 citations, were identified as the most influential authors.
Keywords: Twice Exceptional, giftedness, special education, attention deficit, autism.

INTRODUCTION
The concepts of difference, giftedness, and intelligence, used to describe individuals' various characteristics, abilities, and potential, are among the main topics in fields of knowledge such as psychology, education, and social sciences. While giftedness refers to an individual demonstrating performance significantly above the norm in a particular area compared to others, exceptional talent refers to individuals possessing a noticeably higher potential in certain areas—academic, artistic, and physical—compared to a significant portion of the social group in which they live (Sak, 2014; Sezgin, 2020; Schultz and Schultz, 2007). Giftedness, on the other hand, is defined by neuroscientists and neuropsychologists as having an IQ above average as the fundamental criterion for giftedness (Sak, 2020; Ersoy and Avcı, 2004; Sak, 2014). In addition, the concept of giftedness is used for individuals who demonstrate superior performance beyond what is expected for their age group, possess long-term attention span, can use multiple methods in problem-solving, have long-term and high memory capacity, strong memorization skills, high logical reasoning, can explain facts and events within cause-and-effect relationships, establish analytical and mathematical relationships... etc. (Kaya, 2021; Sak, 2014).

The association of difference with concepts such as superior ability and superior intelligence is interpreted as a one-sided affirmation from the individual's perspective, while research shows that one-sided difference in superior ability or superior intelligence can sometimes be a two-sided characteristic. In other words, it shows that giftedness and superior intelligence in gifted and highly intelligent individuals are unexpected second characteristics in an individual, and that this is a contradictory situation that conflicts with giftedness and superior intelligence. It is now a well-known fact that a successful student in the classroom can also exhibit undesirable characteristics that

lead to failure. In the literature, this is referred to as duality, or in other words, the state of exhibiting both good and not-so-good characteristics, such as superior talent and learning difficulties, attention deficit disorder, or autism. Difference refers to the situation that arises when comparing two objects, while double difference refers to the existence of two opposing characteristics in the same object or individual. In psychology, the concept of dual difference includes having one of the following conditions: special talents, specific learning difficulties, autism spectrum disorder, socio-emotional developmental disorder, attention deficit hyperactivity disorder, and language and speech disorders (Neihart, 2008; Ömür, 2019). Twice-exceptional individuals are defined as a relatively new group in the field of special education. Therefore, it is emphasized that the characteristics of these individuals should be understood and that they should be educated according to their needs (İlker, 2017). The limited number of studies conducted with these individuals in both national and international literature, and the very limited number of studies in Turkey on twice-exceptional individuals with special talents and learning difficulties (Yılmaz et al., 2021), highlight the importance of compiling and organizing publications from internationally renowned journals to draw attention to research in this field. In this context, the bibliometric analysis of academic research on twice exceptionality and giftedness based on the Scopus database has been determined as the objective of this article. To this end, answers to the following questions have been sought:
Scientific studies on "twice exceptionality and giftedness";
1. What is the distribution of article numbers by year?
2. How are they distributed according to academic publication type?
3. How are they distributed according to academic publication language?
4. The most effective articles:
- What is the appearance of their relationships within themselves?
- How do the most effective articles appear in relation to each other?
- What is the distribution of the most effective articles according to citation counts?

5. How are scientific studies on the topic of "double exceptionality and giftedness" distributed according to the most effective source?
6. How are they distributed according to Keywords?
7. How are they distributed according to the most effective authors?

METHOD

The study was designed as a case study, one of the qualitative research designs that allows the researcher to examine the system in depth and address the subject under investigation within its current context by systematically collecting data within a specific system with defined boundaries (Yin, 2009; Chmiliar 2010; Merriam, 2013). In other words, since the aim was to examine the existing situation by analyzing the bibliometric parameters of the study sample obtained from the Scopus database, covering the topics of "Double Exceptionality and Giftedness," it was deemed appropriate to design it as a case study using a qualitative research approach.

Data Analysis

In the analysis of data in this study, bibliometric analysis was used, a research technique that maps big data using parameters such as citations, documents, sources, institutions, countries, authors, and keywords on a specific topic in the field to present the current state of the field, its intellectual structure, and trends in the literature (Passas, 2024; Donthu et al., 2021; Agostini et al., 2020). Bibliometric analysis is conducted using outputs obtained from databases such as Scopus, Web of Science, Google Scholar, etc. The dataset used in the study was obtained from the Scopus database covering the period 1995-2024. The Scopus database was chosen because it is a multidisciplinary database containing the most reliable content from over 7,000 publishers, peer-reviewed scientific articles, books, conference proceedings, and similar research, covering 330 different disciplines (URL-1). The VOSviewer program (version 1.6.20), a current program that enables the accurate and understandable interpretation of large data sets, was used to map the outputs obtained from the database (Van Eck and Waltman, 2014).

FINDINGS

To reveal the relationship between scientific studies indexed in the Scopus database on the topic of "double exceptionality and giftedness," to identify the most cited articles in the field, to determine the most effective sources in the literature, to identify the countries and institutions that have contributed most to the field with their work on the subject, and to present to the reader the evolution of the field through the change in key concepts during the period studied. To this end, a search was conducted in the Scopus database on 11.01.2025 using the "search within" section, searching the article title, abstract, and keywords for the key concepts "twice-exceptional" OR "2e" OR "dual exceptional" OR "multiple exceptional" AND "gifted" were used to search the Scopus database on January 11, 2025. Within the search period covering 1995-2024, a total of 354 documents were identified, including 236 articles, containing the specified key concepts in the titles, abstracts, and keywords preferred by the

authors. Only articles were included in the study. The first study indexed in the Scopus database on this topic was found to be “Multiple exceptionalities: A case study” by Moon, S. M., and Dillon, D. R. (1995).

Findings on Article Counts by Year

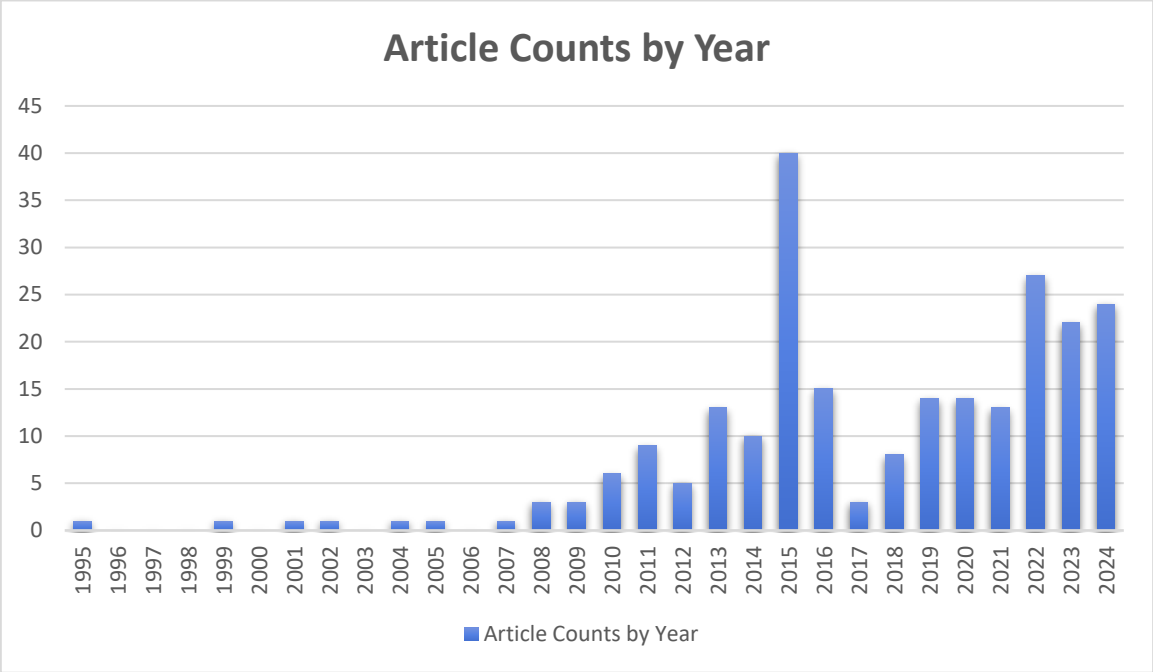


Figure 1. Number of articles by year

When examining the graph of the number of articles per year for the 236 articles on the topic of “Twice Exceptionality and Giftedness” indexed in the Scopus database and covering the period 1995-2024, which constitute the sample of the study, it is seen that the first study was conducted in 1995, but the subject of the study did not feature much in academic studies until 2008. Interest in the subject began in 2008, but until 2013, less than 10 articles per year contributed to the literature. The topic received the most attention in 2015, with 40 articles published, and interest in the topic continued at an inconsistent level until the end of 2024.

Findings on the Distribution of Academic Publication Types

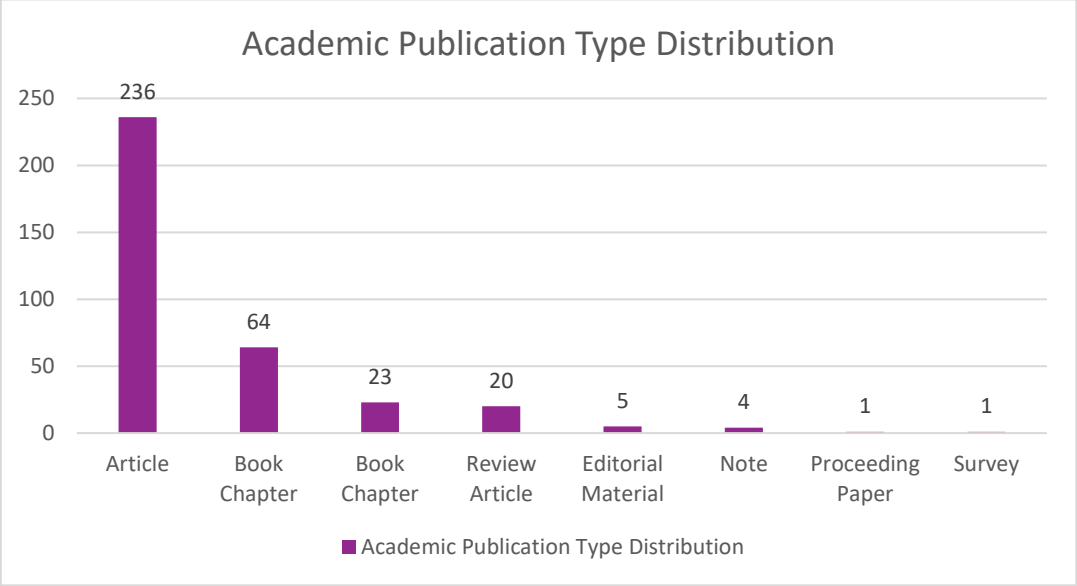


Figure 2. Distribution of academic publication types

On January 11, 2025, a search conducted in the Scopus database using relevant keywords identified 354 documents, of which 236 (66.7%) were articles. It was determined that 64 documents, or 18.1% of all documents, were book chapters, and 23 documents, or 6.5% of all documents, were books.

Findings Regarding the Distribution of Academic Publication Languages

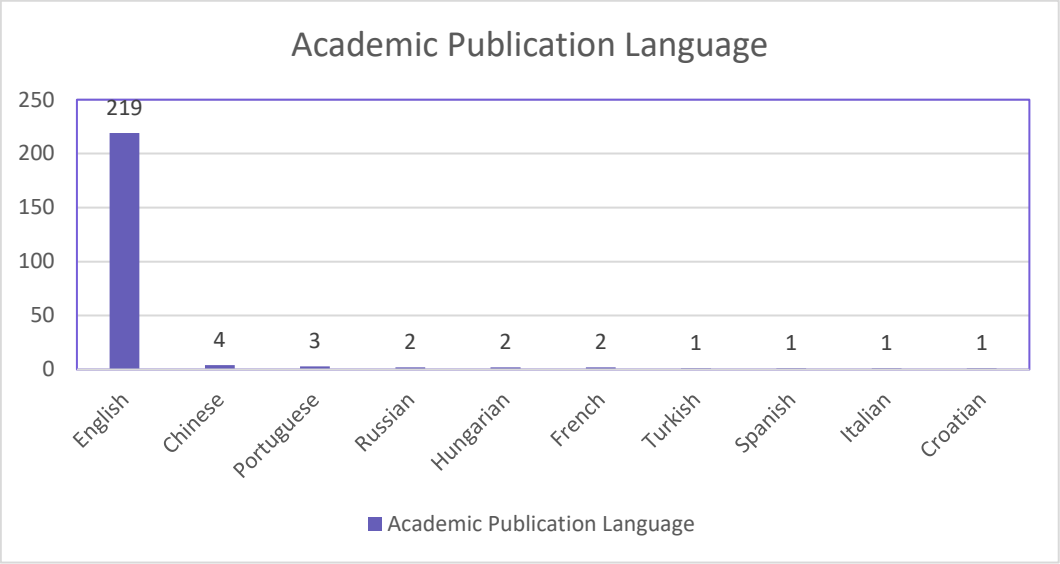


Figure 3. Distribution of academic publication languages

When examining the distribution of publication languages among the 236 articles obtained from the Scopus database and forming the sample of the study, it was determined that 219 articles, or 92.80% of all articles, were written in English. The study also found that research was conducted in Chinese, Portuguese, Russian, Hungarian, French, Turkish, Spanish, Italian, and Croatian. It was determined that the only study indexed in Turkish was the study by Şakar, S. and Köksal, M.S. (2022), "Determining the Opinions of Special Education Teacher Candidates Regarding the Education of Children with Multiple Disabilities."

Findings on the Most Effective Article Analysis

Studies indexed in the Scopus database on the topic of "twice the difference and superior ability" were analyzed using the VOSviewer program to examine citations/documents, presenting researchers with the most cited and effective articles in the relevant literature. In the visual representation, each article is represented as a node within a specific color group. The size of the nodes varies according to the number of citations received by the articles, while the font sizes also differ proportionally to the number of citations received. It is also known that articles in the same color group work on similar topics. Vosviewer visualized the 236 articles (items) comprising the study sample as 80 clusters. Only the first authors of the articles are indicated in the visual.

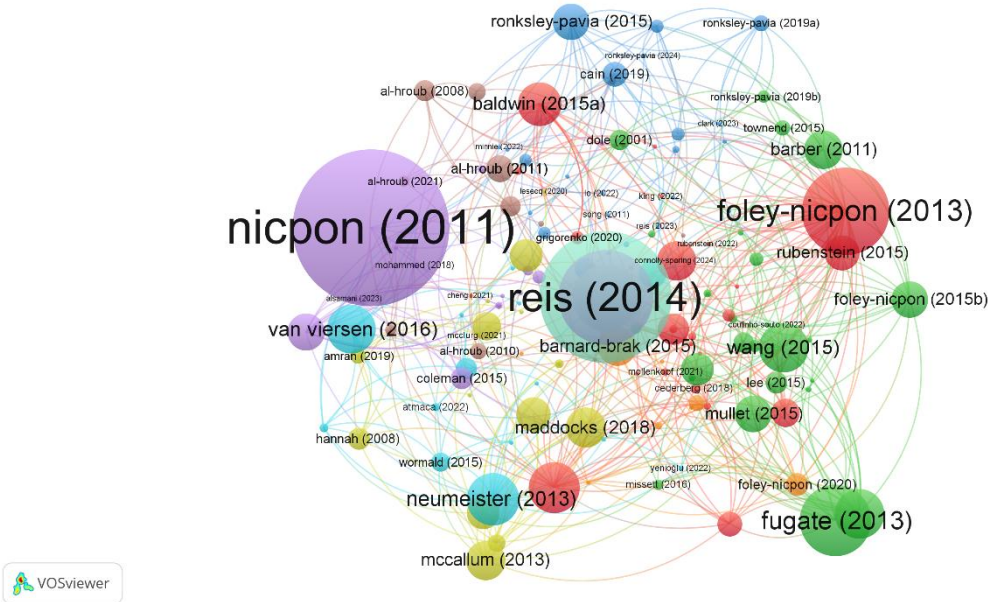


Figure 4. Visualization of the most effective article network map

In the study, the number of articles was reduced to 25, with the condition that each study must have received at least 35 citations, in order to reveal the relationships between articles and identify the most cited articles. The 17 articles (items) identified by the program as having relationships were mapped into 3 clusters.

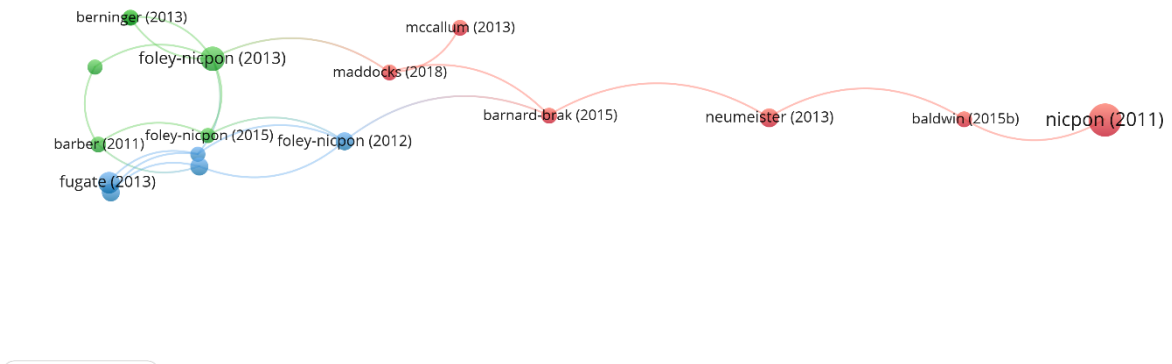


Figure 5. Visualization of the most effective article network map related to each other

The red cluster: Baldwin, L. (2015), Barnard-Brak, L. (2015), Maddocks, D. L. (2018), McCallum, R. S. (2013), Neumeister, K. S. (2013), and Foley Nicpon, M. (2011). The most heavily cited item in the cluster is the study "Empirical investigation of twice-exceptionality: Where have we been and where are we going" by Foley Nicpon, M. and colleagues, which received 188 citations.

The Green Cluster; Barber, C., and Mueller, C. T. (2011), Berninger, V. W., and Abbott, R. D. (2013), Foley-Nicpon, M. (2013), Foley-Nicpon, M., and Assouline, S. G. (2015), Ronksley-Pavia, M. (2015), Rubenstein, L. D. (2015). The most valuable article in the green cluster was found to be the study by Foley-Nicpon et al. (2013), "Twice-exceptional learners: Who needs to know what?", which received 98 citations.

In the third cluster (blue), which includes Foley-Nicpon, M. (2012), Fugate, C. M. (2013), Mullet, D. R., and Rinn, A. N. (2015), Wang, C. W., and Neihart, M. (2015), and Willard-Holt, C. (2013), it was found that the most prominent item in the blue cluster is Fugate, C. M. (2013)'s article titled "Creativity and working memory in gifted students with and without characteristics of attention deficit hyperactive disorder: Lifting the mask", with 78 citations.

The five most cited articles on giftedness and twice-exceptionality have been presented to the reader in Table 1.

Table 1. Most Influential Articles by Citation Coun

First Author	Article Title	Journal / Source	Year	Citation Count
Foley Nicpon, M.	Empirical investigation of twice-exceptionality: Where have we been and where are we going?	*Gifted Child Quarterly*	2011	188
Reis, S. M.	An operational definition of twice-exceptional learners: Implications and applications.	*Gifted Child Quarterly*	2014	155
Assouline, S. G.	Cognitive and psychosocial characteristics of gifted students with written language disability	*Gifted Child Quarterly*	2010	107
Foley Nicpon, M.	Twice-exceptional learners: Who needs to know what?	*Gifted Child Quarterly*	2013	98
Fugate, C. M.	Creativity and working memory in gifted students with and without characteristics of attention deficit hyperactive disorder: Lifting the mask.	*Gifted Child Quarterly*,	2013	78

Upon reviewing the literature on giftedness and twice-exceptionality, it was found that the most highly cited article is Foley Nicpon, M. (2011), with 188 citations. In her study, Foley Nicpon focuses on the identification, support, and educational experiences of twice-exceptional students, offering recommendations for understanding their

needs and providing appropriate educational and support services. Reis, S. M. (2014) ranks second with 155 citations, while Assouline, S. G. (2010), with 107 citations, is identified as the third most influential contribution to the literature.

Findings from the Analysis of the Most Influential Source

A citation/source analysis was conducted to identify the most frequently cited sources within articles focusing on giftedness and twice-exceptionality. In this analysis, each node represents a distinct source, and the size of each node is proportionally mapped according to its citation count. Sources covering similar content areas are visualized within the same color cluster. Based on data retrieved from the Scopus database, it was determined that 97 unique sources were cited across the 236 articles that comprised the study sample.

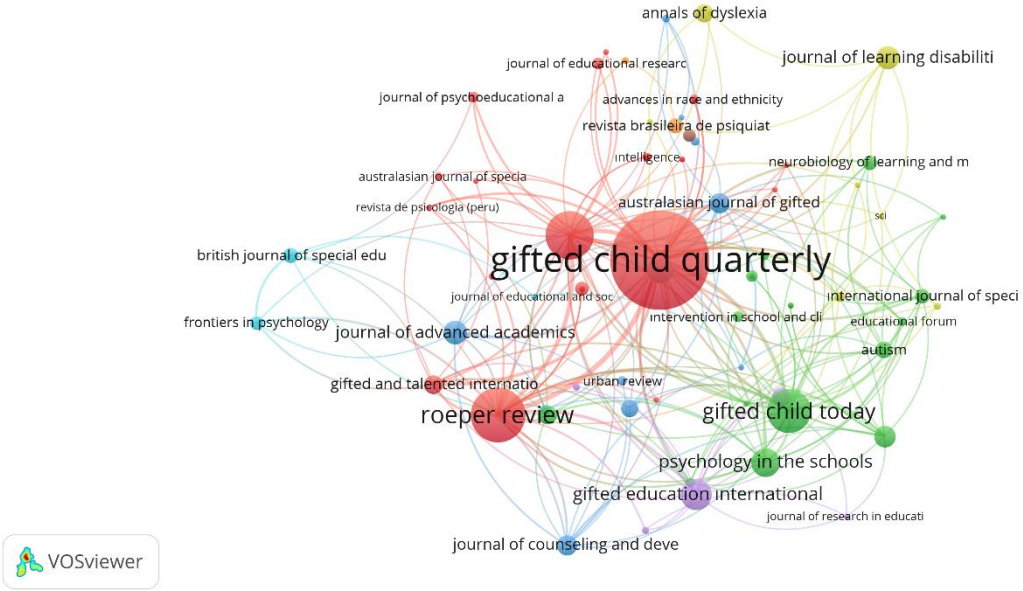


Figure 6. Visualization of the Most Influential Source Network Analysi

To present the most influential sources on twice-exceptionality and giftedness, along with their interconnections and collaborative patterns, a source co-citation analysis was conducted. Among the 97 sources that had published studies on the topic, those meeting the criteria of having authored at least two documents and received a minimum of 10 citations were filtered down to 16 items. These were visualized into three distinct clusters, each representing thematic proximity and citation-based relationships.

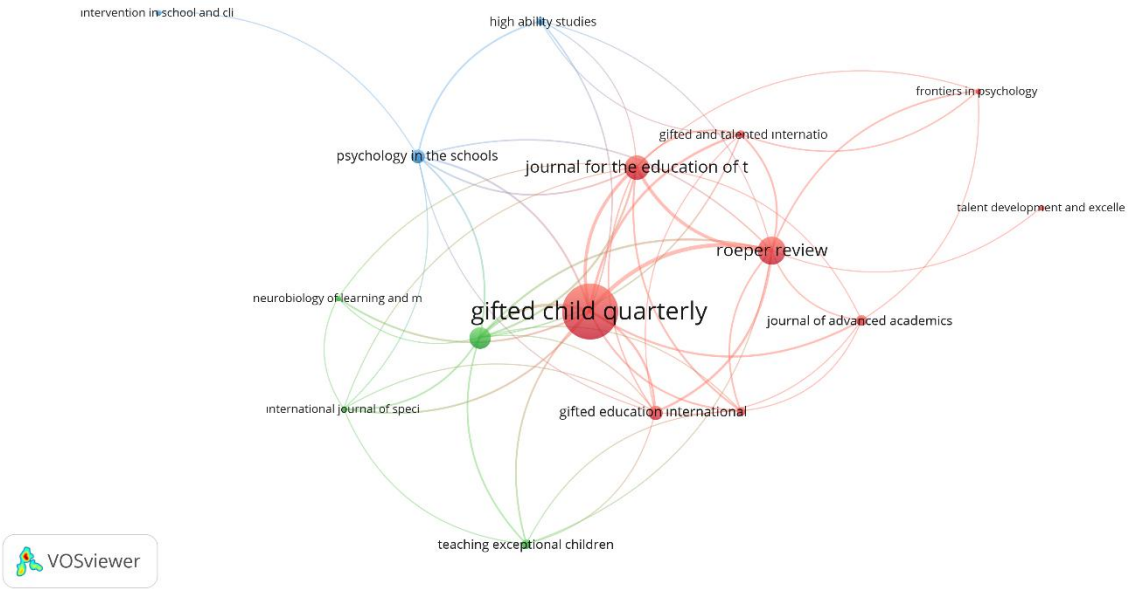


Figure 6. Network Visualization of the Most Influential and Interconnected Sources

It was observed that the red cluster consists of the following sources: Australasian Journal of Gifted Education, Frontiers in Psychology, Gifted and Talented International, Gifted Child Quarterly, Gifted Education International, Journal for the Education of the Gifted, Journal of Advanced Academics, Roeper Review, and Talent Development and Excellence. Within this cluster, Gifted Child Quarterly was identified as the most influential journal, having published 25 documents related to the topic and received a total of 1,157 citations.

The green cluster was found to include the following journals: Gifted Child Today, International Journal of Special Education, Neurobiology of Learning and Memory, and Teaching Exceptional Children. Within this cluster, Gifted Child Today emerged as the dominant source, having published 23 documents on the topic of twice-exceptionality and giftedness, and receiving a total of 200 citations.

The blue cluster was found to include the journals High Ability Studies, Intervention in School and Clinic, and Psychology in the Schools. Within this cluster, Psychology in the Schools emerged as the leading source, having published four documents related to the research topic and receiving a total of 81 citations.

The most influential sources related to the research topic are presented in Table 2 (The table was prepared based on the 2024 updates of journal impact factors).

Table 2. Most Influential Sources by Citation Count

Source Name	**Number of Documents**	**Citation Count**	**Total Link Strength**	**Impact Factor (2024)**
Gifted Child Quarterly	25	1157	74	3.0
Roeper Review	21	314	58	1.7
Journal for the Education of the Gifted	17	244	41	1.2
Gifted Child Today	23	200	30	0.89
Gifted Education International	14	93	19	1.72

Based on citation counts and publication volume, Gifted Child Quarterly was identified as the most influential source in the field, having published 25 documents and received 1,157 citations. With an impact factor of 3.0, the journal publishes original academic research and both quantitative and qualitative studies relevant to the literature (Gifted Child Quarterly, n.d.). Roeper Review ranks second with 21 documents and 314 citations, followed by Journal for the Education of the Gifted, which holds third place with 17 documents and 244 citations.

Findings from the Keyword Analysis

The conceptual structures of articles indexed in the Scopus database on the topic of twice-exceptionality and giftedness were examined to identify the most influential author keywords in the literature. To present the thematic focus and the evolution of research content over time, a co-occurrence analysis of author keywords was conducted. Through keyword analysis, connections were established among the concepts found in article titles, abstracts, and author-defined keywords. This approach enabled the identification of the most frequently used terms within the field, as well as the interpretation of how key concepts have evolved over time. By examining temporal shifts in keyword usage, insights were gained into the changing thematic directions and research priorities in the domain. In the analysis map, each keyword is represented as a node, and both the size of the node and the font of the keyword are proportionally scaled by the software according to the frequency of use—also referred to as the occurrence value. Keywords that are similar or conceptually related are grouped within the same color cluster, allowing for a visual representation of thematic proximity and conceptual relationships.

In the sample of 236 articles retrieved from the Scopus database, a total of 570 distinct keywords were identified. Upon closer examination, it was observed that several variations of the same conceptual term—such as twice exceptional, twice exceptionality, twice exceptionals, twice-exceptional, twice-exceptional (2e), twiceexceptionality, twiceexceptional–dual exceptional, dual exceptionality, dual exceptionalities, and dual-exceptionality—were used inconsistently across different studies. To ensure semantic clarity and analytical precision, a data cleaning process was conducted. During this process, all variations of the term were consolidated under a unified keyword label, allowing for accurate measurement of occurrence frequency and improved conceptual mapping within the keyword analysis.

After data cleaning, 570 keywords were reduced to 508 conceptual terms and mapped into 22 clusters.

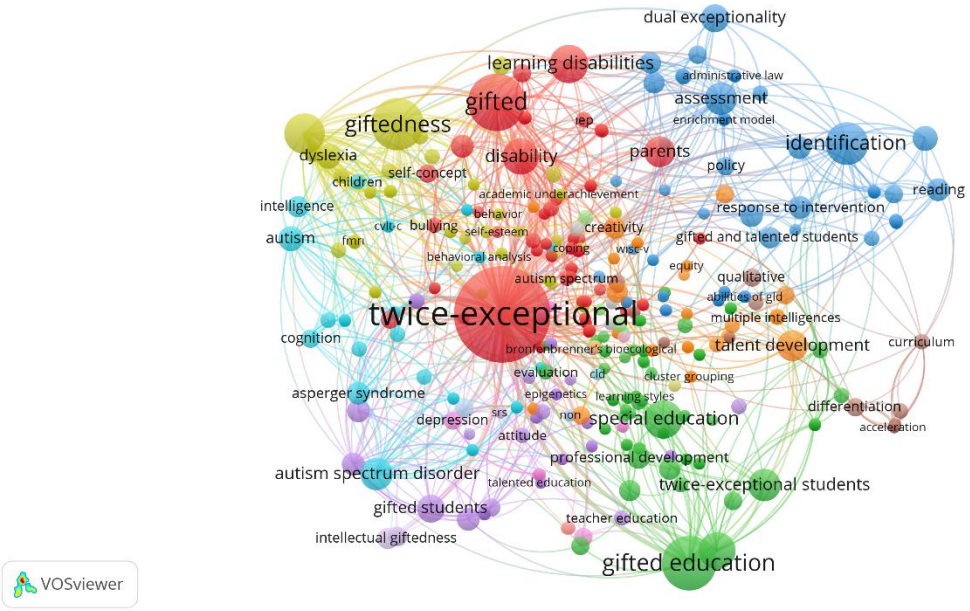


Figure 7. Keyword Network Analysis Visualization

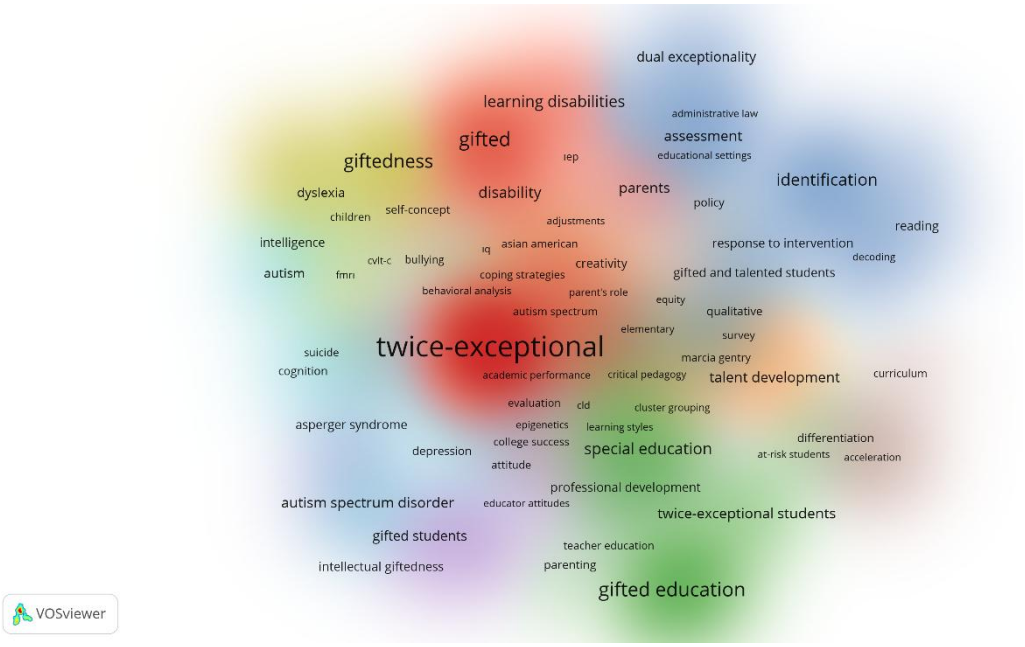


Figure 8. Keyword Density Map Visualization

In the field of twice exceptionality and giftedness research, in order to identify the most frequently used keywords and determine the main areas of focus, the most prominent keywords were reduced to 30 based on a minimum occurrence threshold of five repetitions per concept, and visualized by the software into four clusters.

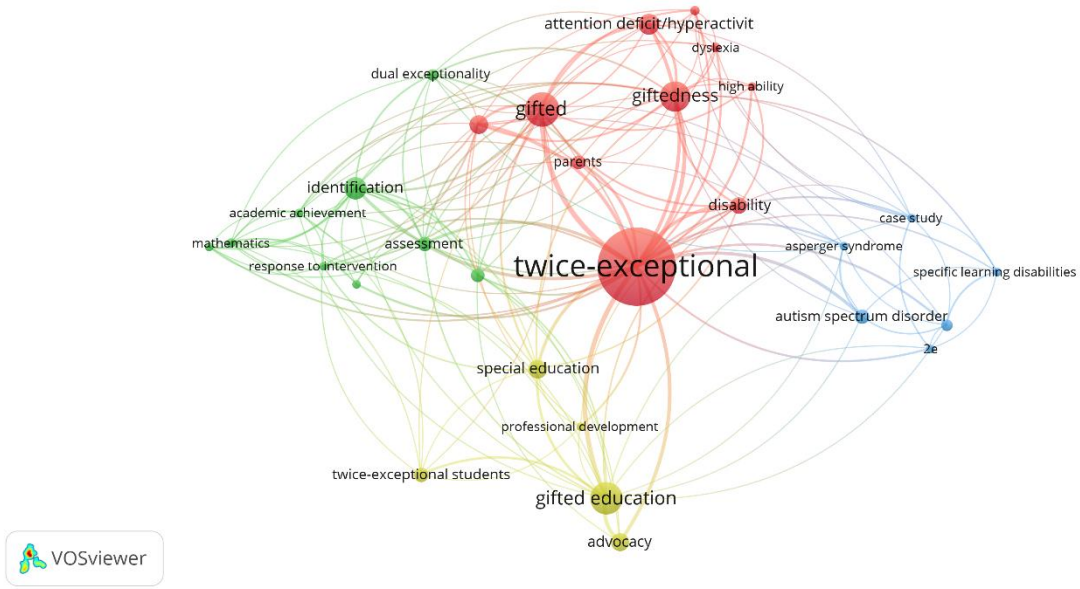


Figure 9. Network Analysis of the Most Prominent Keywords

Cluster 1 (red) was found to consist of the following key concepts: attention deficit/hyperactivity disorder, autism, disability, dyslexia, gifted, giftedness, high ability, learning disabilities, parents, and twice-exceptional. The most dominant keyword in the red cluster was identified as twice-exceptional, with 145 co-occurrence links. In the study by Kanz et al. (2024), titled "Twice-exceptionality unmasked: A systematic narrative review of the literature on identifying dyslexia in the gifted child", the keywords twice-exceptional, dyslexia, and gifted were used.

Cluster 2 (green) was found to include the following key concepts: academic achievement, assessment, dual exceptionality, identification, mathematics, reading, response to intervention, special populations/underserved gifted, and talent development. The most prominent item in Cluster 2, represented in green, was identified as identification, with 21 occurrence links. In the study by Al-Hroub (2013), titled "A multidimensional model for the identification of dual-exceptional learners", the keywords identification, dual exceptionality, and assessment were found to be included.

Cluster 3 (blue) was found to include the following key concepts: 2e, Asperger syndrome, autism spectrum disorder, case study, gifted students, and specific learning disabilities. The most prominent concept in the blue cluster was identified as autism spectrum disorder, with 11 occurrence links. In the article by Burger-Veltmeijer et al. (2016), titled "Intellectually gifted students with possible characteristics of ASD: a multiple case study of psycho-educational assessment practices", the keywords autism spectrum disorder, gifted students, and case study were found to be included.

Cluster 4 (yellow) was found to include the following key concepts: advocacy, gifted education, professional development, special education, and twice-exceptional students. The most prominent keyword in this cluster was identified as gifted education, with 38 occurrence links. In the study by Guo and Chen (2024), titled "Process of policymaking and action to strengthen the discovering of students with twice-exceptionality", the keywords gifted education, special education, and advocacy were found to be included.

To identify the temporal evolution of the most prominent keywords in the literature, the keywords with the highest co-occurrence strength in the field were presented to the reader through overlay visualization analysis.

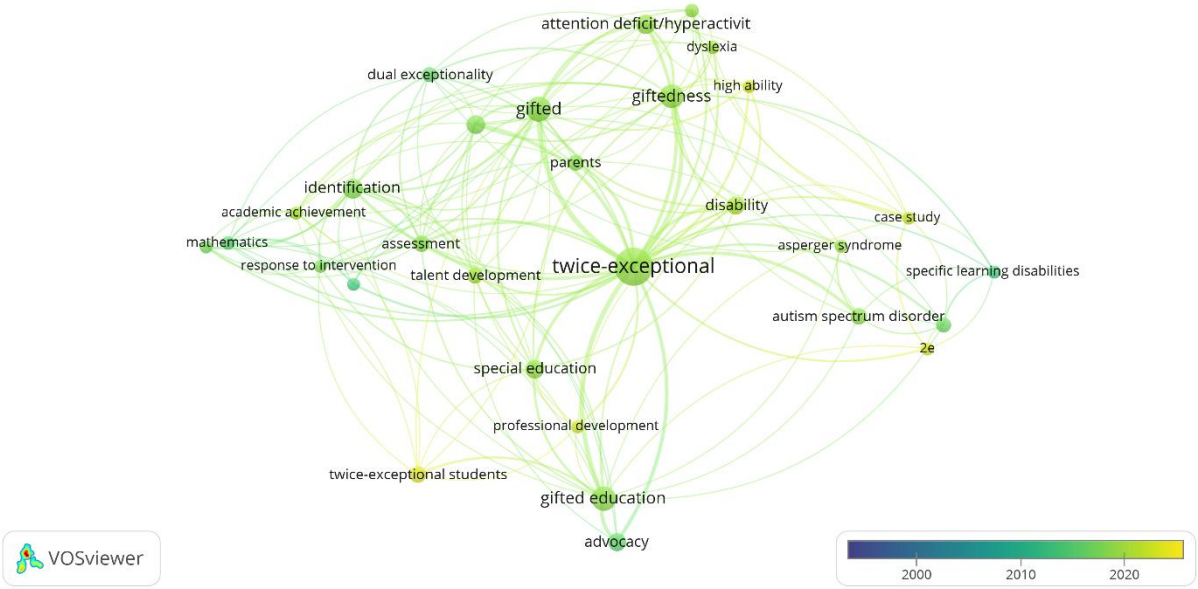


Figure 10. Keyword Overlay Analysis Visualization

The analysis revealed the temporal evolution of key concepts as follows: specific learning disabilities (2013.00), special populations/underserved gifted (2013.20), mathematics (2014.50), advocacy (2015.00), dual exceptionality (2015.38), gifted students (2015.88), reading (2016.40), response to intervention (2017.00), learning disabilities (2017.19), identification (2017.19), assessment (2017.45), parents (2017.60), gifted education (2017.61), autism spectrum disorder (2017.64), gifted (2017.71), autism (2018.00), twice-exceptional (2018.01), giftedness (2018.12), special education (2018.12), attention deficit/hyperactivity disorder (2018.16), Asperger syndrome (2018.20), dyslexia (2018.67), talent development (2018.80), academic achievement (2019.00), disability (2019.29), case study (2020.80), professional development (2021.40), high ability (2021.60), 2e (2022.00), and twice-exceptional students (2022.09). The ten keywords with the highest co-occurrence strength in the field are presented to the reader in Table 3.

Table 3. Top 10 Keywords with the Highest Co-occurrence Strength

Anahtar Kavram	**Keyword**	**Total Link Strength**	**Average Publication Year**	**Average Citation Count**
twice-exceptional	145	608	2018.01	14.31
gifted	42	200	2017.71	15.45
giftedness	34	140	2018.12	9.44
gifted education	38	139	2017.61	8.42
identification	21	112	2017.19	25.90
attention deficit/hyperactivity disorder	19	87	2018.16	14.11
learning disabilities	16	70	2017.19	17.00
disability	14	64	2019.29	8.43
special education	16	59	2018.12	7.94
assessment	11	56	2017.45	35.18

An analysis of the most influential keywords in the literature on twice-exceptionality and giftedness revealed that the keyword “twice-exceptional” ranked first, with 145 co-occurrences and a total link strength of 608. The average publication year for “twice-exceptional” was identified as 2018.01, with an average citation count of 14.31. The second most influential keyword in the study was “gifted”, with 42 co-occurrences and a total link strength of 200, followed by “giftedness” as the third most influential keyword, with 34 co-occurrences and a total link strength of 140.

Findings on the Most Influential Author Analysis

A citation/authors analysis was conducted to identify the most influential authors in the specific literature on twice-exceptionality and giftedness. In the analysis, each author is represented by a node and a color. Authors with

higher citation counts are visualized with larger nodes, while those working on similar topics in their publications are mapped within the same color group.

Within the scope of the analysis, it was found that the 236 articles forming the sample of the study in the Scopus database were authored by 505 individuals. Each of these 505 authors was examined individually, and it was observed that 27 authors were indexed in the database under different name variations such as foley nicpon, megan / foley-nicpon, megan / foley-nicpon, m. / nicpon, megan foley / reis, sally / reis, sally m., etc. Through data cleaning, these variations were consolidated under unified author names. Following the data cleaning process, the number of authors was reduced from 505 to 474 and subsequently mapped.

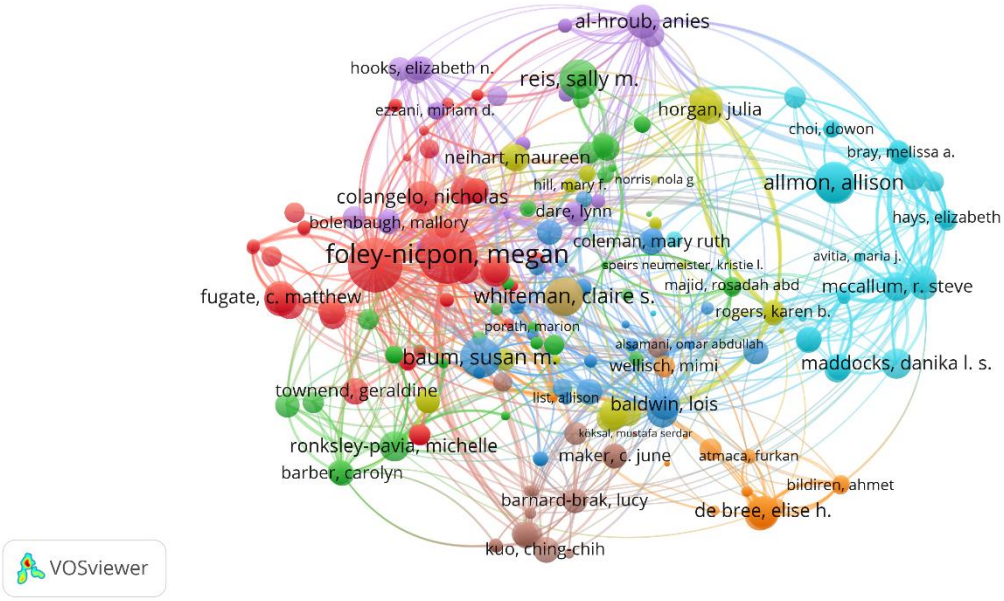


Figure 11. Visualization Map of the Most Influential Author Network

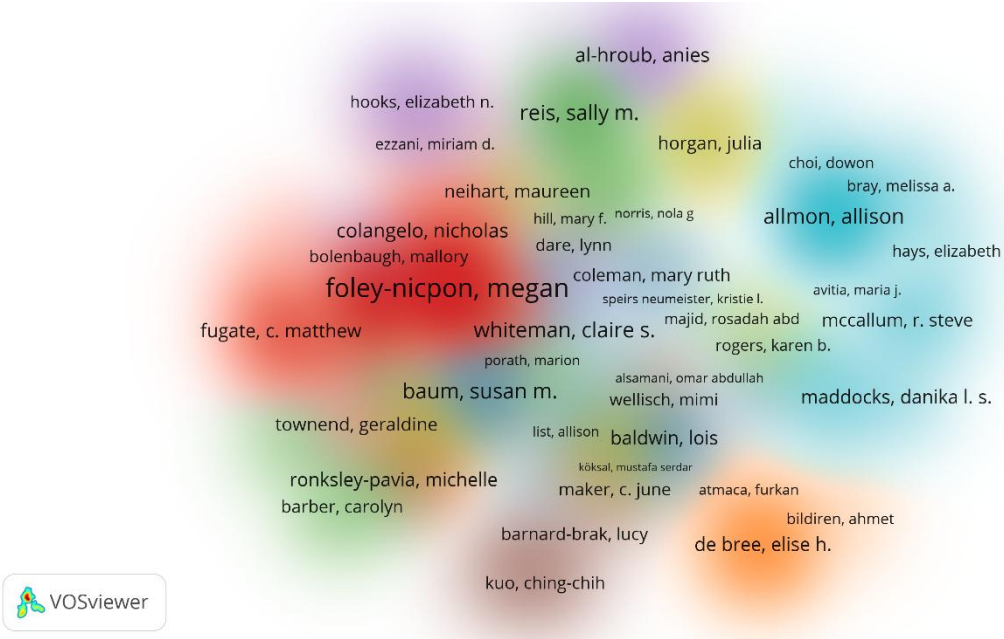


Figure 12. Density Visualization of the Most Influential Authors

To identify the most interconnected influential authors within the scope of the study, authors were filtered based on the criteria of having published at least three documents and received a minimum of ten citations. As a result, the dataset was reduced to 23 authors (items), which were visualized into three clusters.

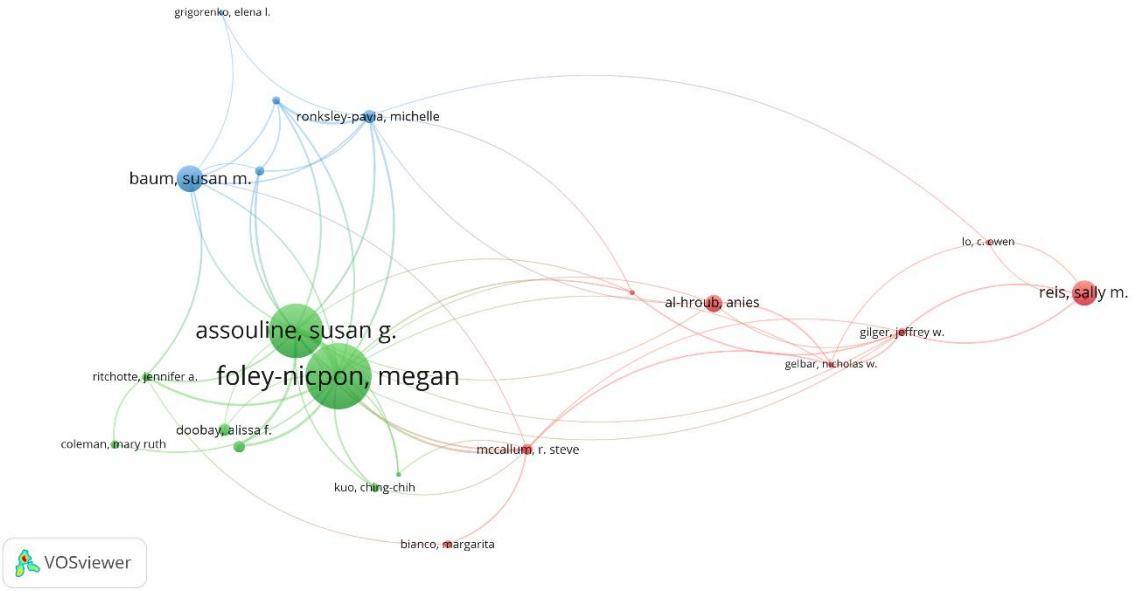


Figure 13. Network Visualization of the Most Interconnected Influential Authors

In Cluster 1 (red), the following authors were identified: Al-Hroub, Anies; Bianco, Margarita; Gelbar, Nicholas W.; Gilger, Jeffy W.; Lo, C. Owen; Madaus, Joseph W.; McCallium, R. Steve; Nickl-Jockschat, Thomas; and Reis, Sally M. Among them, the most influential author in the cluster was determined to be Reis, Sally M., who published six documents and received 188 citations.

In Cluster 2 (green), the following authors were identified: Assouline, Susan G.; Coleman, Mary Ruth; Doobay, Alissa F.; Foley-Nicpon, Megan; Jolly, Jennifer L.; Kuo, Ching-Chih; Mayes, Renae D.; and Richotte, Jennifer A. Among them, the most influential author in the cluster was determined to be Foley-Nicpon, Megan, with 19 documents and 723 citations.

Cluster 3 (blue) was found to include the following authors: Baum, Susan M.; Grigorenko, Elena L.; Pendergast, Donna; Ronksley-Pavia, Michelle; and Townend, Geraldine. Within this cluster, Baum, Susan M. was identified as the most prominent node, having published four documents and received 205 citations.

The top five most influential authors based on their citation counts are presented to the reader in Table 4.

Table 4. Most Influential Authors by Citation Count

Author	Citation Count	Average Citation Count	Average Publication Year	Number of Documents	Total Link Strength
Foley- Nicpon, Megan.	723	38.05	2016.00	19	74
Assouline, Susan G.	553	39.50	2015.14	14	64
Baum, Susan M.	205	51.25	2018.25	4	14
Reis, Sally M.	188	31.33	2018.17	6	3
Al-Hroub, Anies.	112	16.00	2014.86	7	6
Ronksley-Pavia, Michelle.	75	15.00	2019.40	5	22

Within the scope of the study on twice-exceptionality and giftedness, Foley-Nicpon, Megan was identified as the most influential author, having published 19 documents and received 723 citations. Assouline, Susan G. ranked second with 14 documents and 553 citations, while Baum, Susan M. was found to be in third place with 4 documents and 205 citations.

CONCLUSION

As a result of the search conducted in the Scopus database on January 11, 2025, using the keywords "twice-exceptional" OR "2e" OR "dual exceptional" OR "multiple exceptional" AND "gifted", a total of 354 documents were retrieved within the time frame of 1995–2024, 236 of which were classified as articles. The first study on the subject was identified as “Multiple exceptionalities: A case study” authored by Moon, S. M., and Dillon, D.

R. in 1995. The year with the highest number of indexed studies on twice-exceptionality and giftedness in the Scopus database was 2015, with 40 articles. It was found that among the 354 documents identified through the search, 66.7% (236 documents) were articles, 18.1% (64 documents) were book chapters, and 6.5% (23 documents) were books. Of the 236 articles forming the sample of the study, 92.80% (219 documents) were written in English. The study also revealed that publications were available in Chinese, Portuguese, Russian, Hungarian, French, Turkish, Spanish, Italian, and Croatian. It was determined that the only study indexed in Turkish was conducted by Şakar, S. and Köksal, M.S. (2022), titled "Determining the Opinions of Special Education Teacher Candidates Regarding the Course on the Education of Twice-Exceptional Children."

An analysis of the most cited articles in the study revealed that the top-ranked publication was authored by Foley Nicpon, M. (2011) with 188 citations, followed by Reis, S. M. (2014) with 155 citations, and Assouline, S. G. (2010) in third place with 107 citations. The most influential sources publishing on the subject, based on citation counts, were identified as follows: *Gifted Child Quarterly* with 25 documents and 1,157 citations, *Roeper Review* with 21 documents and 314 citations, and *Journal for the Education of the Gifted* with 17 documents and 244 citations.

An analysis of the most influential keywords in the literature on twice-exceptionality and giftedness revealed that the most prominent term was "twice-exceptional," with 145 co-occurrences and a total link strength of 608. This was followed by "gifted," with 42 co-occurrences and a link strength of 200, and "giftedness," with 34 co-occurrences and a link strength of 140.

Within the scope of the study, the most influential authors were identified as Foley-Nicpon, Megan with 19 documents and 723 citations; Assouline, Susan G. with 14 documents and 553 citations; and Baum, Susan M. with 4 documents and 205 citations.

REFERENCES

Al-Hroub, A. (2013). A multidimensional model for the identification of dual-exceptional learners. *Gifted and Talented International*, *28*(1-2), 51-69.

Baldwin, L., Baum, S., Pereles, D., & Hughes, C. (2015). Twice-exceptional learners: The journey toward a shared vision. *Gifted Child Today*, *38*(4), 206-214.

Barber, C., & Mueller, C. T. (2011). Social and self-perceptions of adolescents identified as gifted, learning disabled, and twice-exceptional. *Roeper Review*, *33*(2), 109-120.

Barnard-Brak, L., Johnsen, S. K., Pond Hannig, A., & Wei, T. (2015). The incidence of potentially gifted students within a special education population. *Roeper Review*, *37*(2), 74-83.

Berninger, V. W., & Abbott, R. D. (2013). Differences between children with dyslexia who are and are not gifted in verbal reasoning. *Gifted Child Quarterly*, *57*(4), 223-233.

Burger-Veltmeijer, A. E., Minnaert, A. E., & Van den Bosch, E. J. (2016). Intellectually gifted students with possible characteristics of ASD: a multiple case study of psycho-educational assessment practices. *European Journal of Special Needs Education*, *31*(1), 76-95.

Chmiliar, l. (2010). Multiple-case designs. In A. J. Mills, G. Eurepas & E. Wiebe (Eds.), Encyclopedia of case study research (pp 582-583). USA: SAGE Publications.

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.

Ersoy, Ö. ve Avcı, N. (2004). Üstün zekâlı ve üstün yetenekliler. İstanbul: Çocuk Vakfı Yayınları.

Foley Nicpon, M., Allmon, A., Sieck, B., & Stinson, R. D. (2011). Empirical investigation of twice-exceptionality: Where have we been and where are we going?. *Gifted child quarterly*, *55*(1), 3-17.

Foley-Nicpon, M., Rickels, H., Assouline, S. G., & Richards, A. (2012). Self-esteem and self-concept examination among gifted students with ADHD. *Journal for the Education of the Gifted*, *35*(3), 220-240.

Foley-Nicpon, M., Assouline, S. G., & Colangelo, N. (2013). Twice-exceptional learners: Who needs to know what?. *Gifted Child Quarterly*, *57*(3), 169-180.

Foley-Nicpon, M., & Assouline, S. G. (2015). Counseling considerations for the twice-exceptional client. *Journal of Counseling & Development*, *93*(2), 202-211.

Fugate, C. M., Zentall, S. S., & Gentry, M. (2013). Creativity and working memory in gifted students with and without characteristics of attention deficit hyperactive disorder: Lifting the mask. *Gifted Child Quarterly*, *57*(4), 234-246.

Gifted Child Quarterly (n.d). *Journal overview and metrics. Retrieved January 19, 2025 from https://journals.sagepub.com/overview-metric/GCQ?*

Guo Jingzi, & Chen Yanwei.. (2024). Process of polıcymakıng and actıon to strengthen the dıscoverıng of students wıth twıce-exceptıonalıty. *Contemporary Educational Research Quarterly. 32*(1), 119-121+.

İlker, Ö., & Melekoğlu, M. A. (2017). İlköğretim döneminde özel öğrenme güçlüğü olan öğrencilerin yazma becerilerine ilişkin çalışmaların incelenmesi. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Özel Eğitim Dergisi*, *18*(03), 443-469.
Kaya, F. (2021). Özel yeteneklilerde psikolojik iyi oluş hali. U. Sak (Ed.), Özel yetenekli öğrencilerin sosyal duygusal ve akademik gelişimi içinde (s. 1-21). Pegem Akademi.
Kranz, A. E., Serry, T. A., & Snow, P. C. (2024). Twice-exceptionality unmasked: A systematic narrative review of the literature on identifying dyslexia in the gifted child. *Dyslexia*, *30*(1), e1763.
Maddocks, D. L. (2018). The identification of students who are gifted and have a learning disability: A comparison of different diagnostic criteria. *Gifted Child Quarterly*, *62*(2), 175-192.
McCallum, R. S., Bell, S. M., Coles, J. T., Miller, K. C., Hopkins, M. B., & Hilton-Prillhart, A. (2013). A model for screening twice-exceptional students (gifted with learning disabilities) within a response to intervention paradigm. *Gifted Child Quarterly*, *57*(4), 209-222.
Merriam, S. B. (2013). Nitel araştırma: Desen ve uygulama için bir rehber (3. Baskıdan Çeviri, Çeviri Editörü: S. Turan). Ankara: Nobel Yayın Dağıtım.
Moon, S. M., & Dillon, D. R. (1995). Multiple exceptionalities: A case study. *Journal for the Education of the Gifted*, *18*(2), 111-130.
Mullet, D. R., & Rinn, A. N. (2015). Giftedness and ADHD: Identification, misdiagnosis, and dual diagnosis. *Roeper Review*, *37*(4), 195-207.
Neihart, M. (2008). Identifying and providing services to twice exceptional children. In *Handbook of giftedness in children: Psychoeducational theory, research, and best practices* (pp. 115-137). Boston, MA: Springer US.
Neumeister, K. S., Yssel, N., & Burney, V. H. (2013). The influence of primary caregivers in fostering success in twice-exceptional children. *Gifted Child Quarterly*, *57*(4), 263-274.
Ömür, N. (2019). *Exploration of characteristics, behaviors andeducational needs of children with gifted autism spectrum disorder through grounded theory* (Doctoral dissertation, Master's thesis, Anadolu University]. https://tez. yok. gov. tr/UlusalTezMerkezi).
Passas, I. (2024). Bibliometric analysis: the main steps. Encyclopedia, 4(2).
Ronksley-Pavia, M. (2015). A model of twice-exceptionality: Explaining and defining the apparent paradoxical combination of disability and giftedness in childhood. *Journal for the Education of the Gifted*, *38*(3), 318-340.
Rubenstein, L. D., Schelling, N., Wilczynski, S. M., & Hooks, E. N. (2015). Lived experiences of parents of gifted students with autism spectrum disorder: The struggle to find appropriate educational experiences. *Gifted Child Quarterly*, *59*(4), 283-298.
Sak, U. (2014). Üstün zekalılar özellikleri, tanılanmaları, eğitimleri (4. baskı). Vize Yayıncılık.

Sak, U. (2020). Üstün zekâlılar. Ankara: Vize Akademik Yayınları.
Schultz, D. P. ve Schultz, S. E. (2007). Modern psikoloji tarihi. İstanbul: Kaknüs Yayınları
Sezgin, O. (2020). Psikolojide Kavramlar ve Üstün/Özel (Zekâ, Potansiyel, Yetenek) Kavramına Psiko-Felsefî Açıdan Bakış. Çocuk ve Medeniyet, 5(10), 319-335.
Şakar, S., & Köksal, M. S. (2022). Özel eğitim öğretmen adaylarının iki kere farklı çocukların eğitimi dersine ilişkin görüşlerinin belirlenmesi. *Journal of National Education/Millî Eğitim Dergisi*, *51*(236).
Van Eck, N. J., & Waltman, L. (2014). Visualizing bibliometric networks. In *Measuring scholarly impact: Methods and practice* (pp. 285-320). Cham: Springer International Publishing.
Wang, C. W., & Neihart, M. (2015). Academic self-concept and academic self-efficacy: Self-beliefs enable academic achievement of twice-exceptional students. *Roeper Review*, *37*(2), 63-73.
Willard-Holt, C., Weber, J., Morrison, K. L., & Horgan, J. (2013). Twice-exceptional learners' perspectives on effective learning strategies. *Gifted Child Quarterly*, *57*(4), 247-262.
Yılmaz, H., & Yalçın, H. (2021). *A new-generation parental attitude affecting gifted adolescents' psychological resilience: helicopter parenting* (Doctoral dissertation, Genç Bilge Yayıncılık).
Yin, R. K. (2009). *Case study research: Design and methods* (Vol. 5). sage.
URL-1: https://www.elsevier.com/products/scopus/content son erişim tarihi: 08.01.2025

Social Class, Cultural Capital, and Education: Japanese and International Students' Perceptions of "Oya Gacha"

Yuki AMAKI
College of Economics, Aoyama Gakuin University, Tokyo, Japan
yukiamaki@gmail.com, ORCID: 0000-0002-6794-8870

ABSTRACT
This study examines the perceptions of Japanese and international students in Japan regarding the concept of "Oya Gacha" ("Parent Lottery"). The analysis applied Bourdieu's theory of cultural reproduction as a theoretical framework. "Oya Gacha" reflects the notion that a child's life is significantly influenced by their family's socioeconomic status because it affects their access to educational opportunities and career options. This study used a mixed-methods design incorporating qualitative and quantitative methods to analyze the survey responses of 115 college students. Quantitative analysis indicated that Japanese and international students showed similar overall levels of agreement with the concept, although international students displayed greater variation in their responses. The results suggest that, although both groups recognize the influence of parental background on life outcomes, Japanese students are more likely to agree with the concept of "Oya Gacha." In contrast, international students' perspectives varied based on the social mobility structures of their home countries. Further qualitative analysis revealed that students interpreted the concept differently. Some emphasized structural inequality, whereas others stressed the importance of individual effort in overcoming disadvantaged circumstances. This study highlights the importance of cultural capital in shaping students' beliefs about inequality. The findings contribute to discussions on social inequality by showing how students in different cultural contexts interpret the relationship between family background and individual effort. Future studies should use larger samples and analyze international differences in greater detail.
Keywords: Oya Gacha (parent lottery), educational inequality, cultural reproduction theory, social mobility, cultural capital

1. INTRODUCTION

Over the past few decades, several North American studies have shown that a child's academic achievement and college choices are influenced by their family's socioeconomic status (SES). This includes factors such as parents' income, educational background, and home environment. In the 1970s, French sociologist Pierre Bourdieu introduced the concept of "cultural reproduction" (Bourdieu & Passeron, 1977), which posits that cultural elements of the home, such as parents' knowledge, values, education, and habits, are passed on to children and reinforced through the education system. This leads to the reproduction of social class and highlights how differences in educational and career opportunities are transmitted across generations. This concept has strongly influenced research in sociology and education.

Since around 2020, the term "Oya Gacha" ("parent lottery") has been widely discussed in the Japanese media. It refers to the idea that one's life is largely shaped by the circumstances of one's birth. "Oya Gacha" highlights that children have no control over who their parents are, likening life to a game of chance. It combines "Oya" (parent) and "Gacha" (a random capsule from a toy vending machine or mobile game) to symbolize the randomness of birth circumstances. The term has sparked substantial public debate because it redefines long-standing concerns regarding social inequality in a clear and modern manner. Its circulation through social and mainstream media has raised public awareness of educational inequality and social mobility. Consequently, issues such as educational inequality and social mobility are being widely discussed. In 2021, the Japanese publisher Jiyu Kokuminsha (2021) identified the term as one of the top 10 buzzwords of the year.

Opinions regarding the validity of this concept are divided. Some believe that family background significantly affects educational attainment and success. Hida (2021) interviewed Professor John List of the University of Chicago about the "parent lottery," and he concurred with the concept, emphasizing that parents' wealth and SES strongly influence their children's academic performance. He argued that because parents shape both genetics and the environment, the "parent lottery" is a valid concept. Furthermore, several studies have shown that financial stability and parental education are important for accessing educational and career development opportunities.

In contrast, some view "Oya Gacha" as less valid because it ignores personal effort and negates responsibility. They argue that success is determined not only by parents but also by individuals' hard work and good choices.

However, the interview by Hida (2021), Professor List refuted the claim that the "parent lottery" denies personal responsibility, noting that not all wealthy children succeed, and that some individuals overcome disadvantages. Thus, this study examined how Japanese and international college students in Japan perceive "Oya Gacha," and explored how cultural and social contexts shape their views.

2. RESEARCH QUESTIONS

This study analyzed how Japanese and international students view "Oya Gacha" in Japan, based on the perspective of cultural reproduction theory. This study also examines how students' environments and social contexts influence their opinions on "Oya Gacha." The argument put forth in this study is that students' cultural backgrounds and SES likely shape their perceptions of "Oya Gacha." Japanese students, especially those from disadvantaged backgrounds who are more familiar with the term, often recognize the effects of inequality. However, international students' perspectives may vary according to their home countries' social systems and values. Those from societies that emphasize personal effort may be more skeptical of the concept, believing that individual choices determine life outcomes.

3. THEORETICAL PERSPECTIVE

3.1 BOURDIEU'S THEORY OF CULTURAL REPRODUCTION

"Oya Gacha" can be analyzed through the lens of Bourdieu's theory of cultural reproduction. Bourdieu and Passeron (1977) argued that family, education, and social class shape people's thoughts and behaviors. These factors transcend personal opinions or values and become a shared social structure that people subconsciously internalize (Bourdieu & Passeron, 1977). This structure allows people to subconsciously follow social rules that facilitate daily life. Thus, ingrained behaviors are learned through family life, schooling, and social class.

3.2 TYPES OF CAPITAL AND THEIR SOCIAL EFFECTS

Recently, the term "Oya Gacha" has been widely discussed in Japan, raising awareness of how the family environment affects a person's life. It is based on the idea that a child's future is largely influenced by the family into which they are born. This concept is closely related to Bourdieu's notion of cultural reproduction. Bourdieu and Passeron (1977) emphasized that capital is not limited to money but exists in various forms, all of which have significant effects on individuals' lives. Although "capital" is often associated with economic wealth, Bourdieu argued that cultural, social, and symbolic capital are crucial in determining social success. Bourdieu and Passeron (1977) argued that societal success is not only determined by individual effort but also strongly influenced by wealth, knowledge, and social connections. Kosutic (2017), who conducted a study of 534 high school seniors, provided empirical evidence to support Bourdieu's theory of cultural reproduction. Privileged students tend to excel academically and attend college at higher rates than their less-privileged peers (Kosutic, 2017). Thus, a person's birth circumstances can significantly affect their life trajectory, and this inequality is perpetuated and reproduced across generations, thereby deepening social disparities. Moreover, these forms of capital vary according to family background and social class. For example, children from affluent families tend to have greater access to quality education, and their social connections and status often provide them with significant advantages in terms of career opportunities. Children from low-income families struggle to acquire cultural capital, which limits their educational and career opportunities. The concept of "Oya Gacha" refers to this inequality.

4. LITERATURE REVIEW

4.1 "OYA GACHA" AND THE THEORY OF CULTURAL REPRODUCTION

"Oya Gacha" refers to the idea that a child's life is significantly influenced by the circumstances of their birth. This concept is closely related to Bourdieu's theory of cultural reproduction (Bourdieu & Passeron, 1977). High-SES children adjust more easily to school, whereas low-SES children struggle more, thus widening the academic gap between rich and poor. Sengonul (2022) found that SES shapes the impact of personal commitment on achievement, which reinforces educational inequality. Grant and Kniess (2023) analyzed rural students' experiences at flagship universities in the U.S. and how they adapted and succeeded compared to their urban counterparts. Their findings were consistent with the notion of "Oya Gacha" in Japan, further indicating that a child's future is determined by family environment factors, including parents' incomes and educational backgrounds. Moreover, high-SES children experience greater parental involvement in maintaining their status, whereas low-SES children often receive less parental input. White et al. (2020) analyzed the challenges faced by first-generation and low-SES college students from a cultural capital perspective, highlighting how parental education influences a student's ability to adjust to college life.

4.2 PARENTAL INVOLVEMENT AND ACADEMIC ACHIEVEMENT

Parental involvement at home and school positively affects academic achievement. Sengonul (2022) argued that parental involvement has a particularly strong effect on children from low-income families because educational support at home plays a critical role in academic improvement in economically disadvantaged environments.

However, when parents themselves have limited education or are economically disadvantaged, providing educational support to their children can be challenging (Sengonul, 2022).

Grant and Kniess (2023) found that rural students experience significant cultural differences in terms of language, food culture, and dress compared to urban students, which affects whether cultural habits acquired at home match the university environment. Urban parents can often provide their children with the knowledge and experience required to help them adjust to the college environment, whereas rural parents may struggle to do so. Consequently, rural students face greater challenges, not only academically but also socially and culturally. White et al. (2020) found that first-generation college students struggle to adapt to academic culture, college rules, and academic literacy requirements more than students whose parents attended college. Accordingly, the concept of "Oya Gacha" captures the structural advantage that children from educated involved families have in accessing and succeeding in higher education. In contrast, children without such support often find it difficult to adapt. According to Bourdieu's concept of cultural capital, students who succeed in higher education typically utilize knowledge and skills inherited from their families. However, many first-generation college students lack this advantage, increasing the likelihood of encountering academic and social difficulties (White et al., 2020).

4.3 ROLE OF SOCIOECONOMIC STATUS

SES mediates the impact of parental involvement on academic success. High-SES families invest in education, instill study habits in their children, and utilize strong networks. In contrast, in low-SES families, although increased parental involvement can positively affect children's academic performance, financial constraints and a lack of educational experience may limit its effectiveness. White et al. (2020) analyzed how SES differences affect educational outcomes and college adjustment, highlighting the difficulty low-SES students face in "getting a read on college." Even when they have the opportunity to attend college, financial burdens and cultural differences make it difficult for them to persist in their studies.

Sengonul (2022) also found that children from high-SES families are more likely to engage in intellectual activities, such as visiting museums and reading books, because they understand the value of formal culture. In contrast, children from low-SES families receive fewer opportunities to engage in such cultural activities and are less exposed to informal education. These differences significantly affect academic performance and dropout rates in higher education. Thus, as the concept of "Oya Gacha" suggests, children's lives are heavily influenced by the family environment into which they are born, which is a factor that they cannot control. These prior empirical findings are consistent with Bourdieu's theory of cultural reproduction, which describes how socioeconomic inequalities are perpetuated across generations through unequal access to cultural capital and education.

5. RESEARCH METHODOLOGY

This study focused on Japanese and international university students in Japan. Convenience sampling was used to recruit participants from intercultural studies courses at two Japanese universities. The questionnaire took approximately 20 minutes to complete. Participation in the survey was voluntary, and all respondents provided informed consent before participation. Participants were assigned ID numbers to protect their identities. Participant data pertaining to gender and country of origin were excluded to protect their identities. Data related to country of origin were excluded because of the risk that such information could be used to identify participants, particularly where the number of international students from a certain country was small.

This study employed quantitative and qualitative analyses. In the quantitative components, respondents evaluated the validity of "Oya Gacha" using a 4-point Likert scale ranging from "totally agree" to "totally disagree." In the qualitative items, respondents provided free-text responses explaining their thoughts as to whether and why "Oya Gacha" was valid. Quantitative data were analyzed using SPSS software, and a cross-tabulation analysis was conducted. Descriptive statistics were initially calculated to summarize the response patterns, including the means and standard deviations for all groups. Then, group differences were examined using Welch's t-test because the assumption of equal variances was not met.

Free-text responses were analyzed using KH Coder (Higuchi, 2016, 2017) to identify word patterns and co-occurrence networks. This approach enabled the examination of how key concepts and themes emerged in the participants' responses. First, all free-text responses were anonymized and imported into KH Coder. Variations in spelling and wording were standardized such that words with the same meaning were treated consistently in the analysis. Second, a morphological analysis was conducted to segment the text and identify frequently occurring nouns, verbs, and descriptive expressions. Third, a co-occurrence network was generated to show which words tended to appear together across responses. Fourth, the clusters identified by KH Coder were examined and compared with the original responses to develop broader interpretive themes. This step was necessary because network results alone do not automatically produce themes. Instead, the themes were developed by repeatedly

reading the clustered words and full student responses. Finally, representative quotations were selected to illustrate each theme, ensuring that the interpretations remained closely grounded in the original data. Qualitative analysis followed a two-step procedure to improve analytic transparency. Keyword patterns were identified using a computer program, and then the themes were interpreted. This approach enabled the systematic text analysis to be combined with careful consideration of the contextual meanings conveyed in students' responses.

6. QUANTITATIVE ANALYSIS

A survey was conducted among Japanese students (n = 85) and international students (n = 30) on their perceptions of "Oya Gacha." The survey used a 4-point Likert scale, and the response patterns in the data were analyzed. Table 1-1 presents the descriptive statistics, including the means (M), standard deviations (SD), and coefficients of variation (CV) for both groups. Table 2 presents the response distributions. Table 1-2 reports the results of Welch's t-test comparing the two groups. An independent-samples t-test was conducted to examine whether the two groups demonstrated any significant differences.

Table 1-1: Descriptive statistics of responses on "Oya Gacha" among Japanese and international students

Group	N	Mean	SD	CV
Japanese students	85	3.16	0.61	19.30
International students	30	3.23	0.77	23.84

Table 1-2: Results of Welch's t-test comparing Japanese and international students

t	*df*	*p*	Mean difference	SE difference	95% CI lower	95% CI upper
-0.44	42.60	.663	-0.07	0.16	-0.38	0.25

Table 2: Distribution of responses on "Oya Gacha" among Japanese and international students

		The Concept of "Oya Gacha"				
		Totally disagree	Somewhat disagree	Somewhat agree	Totally agree	Total
Japanese students	N	1	7	54	23	85
	%	1.2%	8.2%	63.5%	27.1%	100%
International students	N	0	6	11	13	30
	%	0%	20.0%	36.7%	43.3%	100%

Tables 1-1 and 1-2 show that the mean score for Japanese students was 3.16 (SD = 0.61), whereas that for international students was 3.23 (SD = 0.77). Japanese and international students showed no significant differences, $t(42.60) = -0.44$, $p = .663$. Welch's t-test was conducted because the assumption of equal variances was not met. The mean difference was −0.07, 95% confidence interval (CI) [−0.38, 0.25]. The effect size was very small (Cohen's $d = -0.10$). These results suggest that both groups generally agree with the concept of "Oya Gacha." However, the SD (0.77) and CV (23.84) for international students were higher than those for Japanese students (SD = 0.61, CV = 19.30). These results suggest that opinions on "Oya Gacha" may vary more widely among international students.

Table 2 shows that 27.1% of Japanese students strongly agreed, while a higher proportion (43.4%) of international students selected the same response, indicating stronger agreement among international students. However, 8.2% of Japanese students chose "somewhat disagree," compared to 20.0% of international students. This suggests that international students were also more likely to view "Oya Gacha" as ambiguous. Overall, these descriptive results indicate that although the average levels of agreement are similar, the distribution of responses among international students is more polarized.

Table 2 further shows that 90.6% of Japanese students combining "somewhat agree" and "totally agree" supported the concept of "Oya Gacha," indicating strong recognition of the intergenerational transmission of SES. Among international students, 43.3% strongly agreed; however, 20.0% somewhat disagreed. This is potentially because acceptance of the concept of "Oya Gacha" may be lower in countries with greater social mobility, whereas students from countries with greater educational inequality may adhere to the idea more.

This study examined differences in perceptions through the lens of cultural reproduction theory, which links families' cultural capital to educational attainment. In Japan, educational opportunities are highly dependent on parental financial status and educational background, indicating that Japanese students may consider the concept

of "Oya Gacha" to be relatable. The results in Table 1 show that although the mean score of Japanese students (3.16) was almost the same as that of international students (3.23), the lower SD and CV indicate that Japanese students' opinions were more consistent. This may be because of the strong presence of the cultural reproduction of inequality in Japan, making awareness of the influence of family background more common and widespread. In contrast, the greater variation in responses among international students may be related to variations in social structures and educational systems between countries.

7. QUALITATIVE ANALYSIS

7.1 JAPANESE STUDENTS' POSITIVE RESPONSES (AGREEMENT WITH "OYA GACHA")

In this study, 115 students shared their views on the validity of the concept of "Oya Gacha." Figure 1 shows a co-occurrence network of key terms extracted from students' responses. KH Coder software was used to analyze characteristic words extracted from positive and negative appraisals made by Japanese and international students. The KH Coder analysis identified three key themes in Japanese students' positive responses (Figure 1).

7.1.1 PARENTAL FINANCE: IMPACT ON CHILDREN'S LIVES

Japanese students observed that parental finances influence educational and career opportunities, positing that wealthier families offer stability, whereas financial struggles limit these opportunities. They argued that "Oya Gacha," highlights how family finances shape a child's future. Japanese students provided the following comments:

> *Financial stability aids success, but money alone doesn't define happiness. Even less wealthy parents can be great.*
>
> *Effort matters, but parental income shapes the future. I could only attend college with my grandparents' financial support, proving "Oya Gacha" exists.*
>
> *Children from low-income families have fewer education options. Many can't reach the starting line despite effort, as parental investment shapes opportunities.*

These Japanese students emphasized that family finances determine education and career options, thus reinforcing "Oya Gacha." These comments suggest that students recognized the influence of economic resources on access to educational and future life opportunities. Several responses indicated that financial support from family members plays a critical role in enabling students to pursue higher education. This perception reflects the notion that family economic capital affects educational trajectories and contributes to the reproduction of social inequality.

7.1.2 THE CONCEPT OF "OYA GACHA" AND INFLUENCE OF THE ENVIRONMENT

Some Japanese students recognized that environmental factors influence outcomes, while also contending that effort can foster success. They viewed "Oya Gacha" as an important social issue among several others and provided the following comments:

> *Children cannot choose their parents. In an education-oriented society, parental wealth shapes the future, but national and environmental factors also matter.*
>
> *"Oya Gacha" is partly true since parents shape children's lives. It also shifts blame from individuals to systemic issues, raising social awareness.*

These comments highlight the perceived importance of parental backgrounds, while acknowledging that effort and wider social support can play key roles in determining life outcomes. These responses also suggest that some students view "Oya Gacha" as a reflection of family background and a means of drawing attention to social and institutional conditions. In this sense, the concept is viewed as raising awareness of the structural factors that influence educational opportunities and life outcomes.

7.1.3 RELATIONSHIP BETWEEN EDUCATION AND POVERTY

Some students focused on how financial status affects education and limits options. They recognized that differences in values across society are influenced by economic disparities, while qualifying this by stating that other factors can limit this effect. These Japanese students provided the following comments:

> *"Oya Gacha" is mostly true. Family wealth determines educational advantages, with wealthy children accessing private schools and children from poor families struggling with tuition.*

"Oya Gacha" is partially true. Parents' values and financial situation influence a child's choices, but children can still find hope through community support, scholarships, and educational programs. Financial aid can help mitigate educational barriers.

These students recognized that poverty limits educational opportunities and future choices, reinforcing and propagating "Oya Gacha." Some responses indicated that external support systems, such as scholarships and community programs, could help reduce the disadvantages associated with low family income. This suggests that students recognize the structural impact of economic inequality and perceive the possibility of reducing educational disparities through institutional support and social policies.

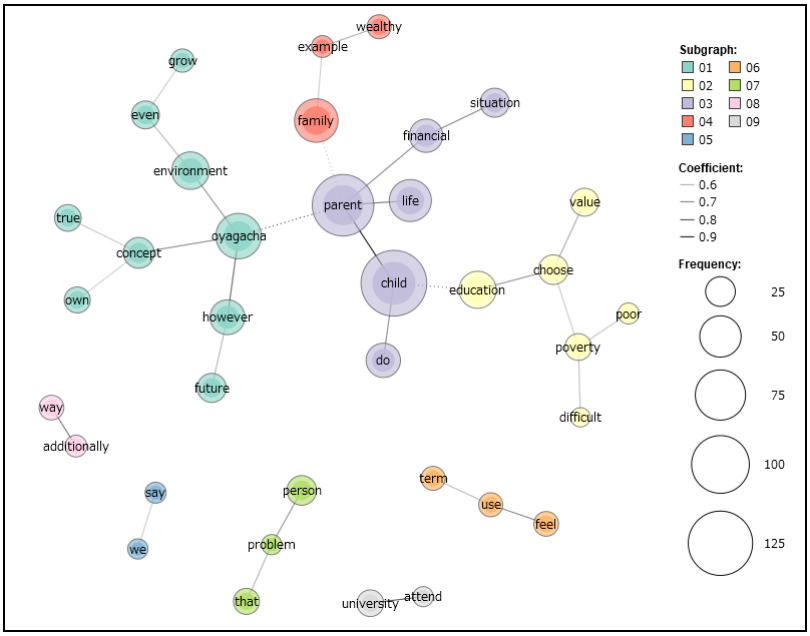


Figure 1: Co-occurrence Network, Japanese Students' Positive Responses

7.2 JAPANESE STUDENTS' CRITICAL RESPONSES (DISAGREEMENT WITH "OYA GACHA")

KH Coder identified recurring terms in Japanese students' critical responses. These opinions can be categorized into three themes (Figure 2).

7.2.1 IMPORTANCE OF PERSONAL EFFORT

These students argued that effort, and not parental influence alone, shapes life outcomes. These students viewed "Oya Gacha" as misleading or extreme. While they acknowledged the existence of parental influence, they rejected the notion that one can blame parents for everything. One student noted, "Having bad parents doesn't mean failure. Life changes with effort. Instead of blaming parents, take responsibility and act." This group contended that individual responsibility and ownership over circumstances are vital determinants of success. Another student stated, "We shouldn't blame parents. Even children from poor families can change their situation with effort." These responses suggest that some students strongly emphasize individual responsibility and personal effort in shaping life outcomes. Although they acknowledge that family background can affect opportunities, they reject the notion that it entirely determines a person's future. From this perspective, the concept of "Oya Gacha" might be considered excessively deterministic because it underestimates the importance of individual agency and perseverance.

7.2.2 THE SOCIAL SYSTEM MATTERS

These students saw broader societal structures, not parents, as the real issue. The presence of words such as "responsibility," "problem," and "shift" indicated their belief that society should not place all the blame on parents, but instead rethink the entire system. Furthermore, the co-occurrence of "overcome," "influence," and "system" suggests that, even in difficult economic situations, improving one's life is possible by utilizing social support systems and initiating effort. One student wrote, "'Oya Gacha' diverts attention from systemic issues, misplacing blame on parents instead of society. It oversimplifies poverty and hinders systemic reform." This group believed that placing the blame on parents ignored the responsibility of the government and wider society for the reproduction of inequality. These responses suggest that some students interpret "Oya Gacha" as a concept that obscures deep structural societal problems. Instead of focusing solely on family background, they emphasize the

roles of social institutions, public policy, and economic structures in shaping educational and life opportunities. From this perspective, discussions of "Oya Gacha" can encourage reflection on the structural causes of inequality and need for systemic reform.

7.2.3 OVERCOMING ADVERSE LIFE CIRCUMSTANCES

Some students argued that despite poverty being a significant factor, life is shaped by personal choices and actions. These students saw life as being determined by choices, not chance. The link between "college" and "poor" denoted how scholarships and effort were seen as helping disadvantaged students access higher education. One student stated, "Many students who are born into poor families study hard, earn scholarships, and attend college. We can't choose our parents, but a good life depends on our actions." Another student wrote, "We cannot choose our parents, but whether we can lead a good life depends on our own actions." A belief that personal choices and effort are more important than the situation into which individuals are born was prevalent among this group. These responses suggest that some students emphasized the importance of individual agency and perseverance in overcoming disadvantaged circumstances. Although they acknowledged the constraints associated with family background, they also highlighted the possibility of upward mobility through effort, education, and institutional support, such as scholarships. This perspective reflects a belief in meritocratic values, in which personal initiative and hard work are considered essential for success.

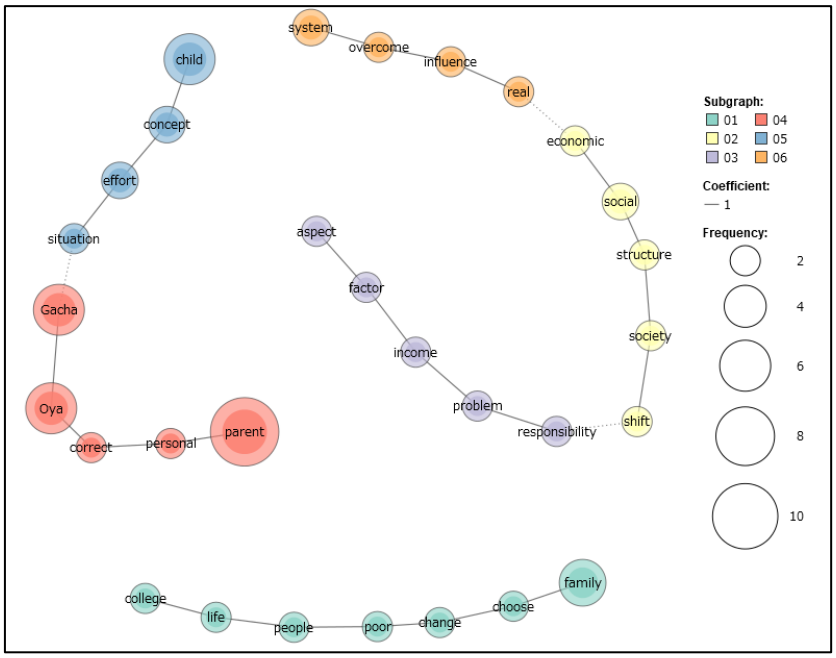


Figure 2: Co-occurrence Network of Japanese Students' Critical Responses

7.3 INTERNATIONAL STUDENTS' POSITIVE RESPONSES (AGREEMENT WITH "OYA GACHA")

The co-occurrence network analysis categorized international students' positive views of "Oya Gacha" into three themes (Figure 3).

7.3.1 THE INFLUENCE OF "OYA GACHA" AND THE FAMILY ENVIRONMENT

This group highlighted that family wealth shapes education and opportunities, with many international students agreeing that financial status is crucial for shaping a child's future. International students stated the following:

> *Wealthy children can access quality education stress-free, while poorer children struggle with financial worries.*

> *A family's financial situation has a great impact on a child's mental state, which affects his or her personality.*

These comments show that some international students felt that parental wealth determines education quality, and economic status affects well-being and the formation of personality traits. Other international students stated the following:

> *People born into poor families have to work much harder than others, especially compared to those*

born into wealthy families, to escape the cycle of poverty.

Differences in family values have a big impact on a child's life.

These responses suggest that many international students recognized the significant impact of the family environment on educational opportunities and life outcomes. In particular, financial resources and family values were viewed as important factors that shape educational access and personal development. This reflects a recognition that family background can create unequal starting points for children in society.

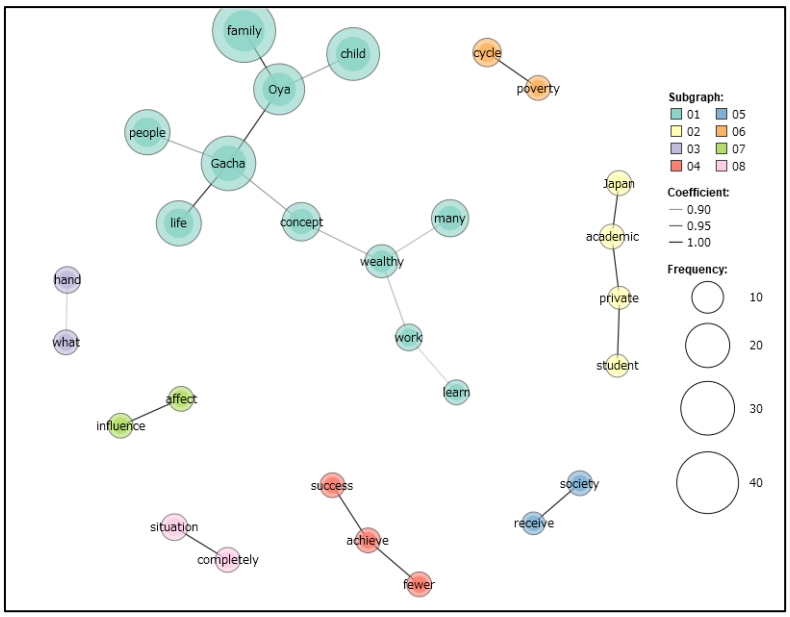


Figure 3: Co-occurrence Network of International Students' Positive Responses

7.3.2 JAPAN'S EDUCATION SYSTEM

This theme pertained to how Japan's education system influences students from different backgrounds. The terms "academic," "private," and "student" suggested that international students recognize this impact. The reference to "private" denotes that some international students believe private schools help secure academic credentials. As "Oya Gacha" relates to opportunity, many international students noted that wealthier families can afford private education, thereby increasing their chances of accessing top universities and acquiring good jobs. One student noted, "Wealthy students attend private schools and take music and sports lessons. For poor children, and even for children from average families, this is something rare." This group highlighted Japan's educational inequality, whereby wealthy children can access private education and extracurricular activities, whereas children from lower-income families lack access to these opportunities.

Some international students compared Japan's educational system to that of their own country.

In South Korea, the term "gold spoon" refers to the idea that a person's academic background and job opportunities are determined by their parents' financial status, affecting their entire life.

These students highlighted Japanese society's strong emphasis on individual effort in adulthood. Their comparisons also suggested that other societies have similar perceptions of family-based inequality, even if they are expressed differently. These observations show that international students understand the concept of "Oya Gacha" in the Japanese context and in relation to patterns of educational inequality in their home countries.

7.3.3 THE DIFFICULTY OF ACHIEVING SUCCESS

This theme highlights how family background affects success. The word "fewer" reflects the belief of some students that low-income individuals have limited opportunities, with the system favoring wealthier students. Some international students commented the following:

Children born into low-income families have fewer opportunities to get a good education compared to those from wealthy families.

Many people use the term "Oya Gacha" as an excuse for their failures in life.

These comments indicate that some students believed that children from low-income families have fewer educational opportunities.

There are many people who have achieved success despite growing up in a difficult environment.

People from low-income families worked twice as hard and overcame challenges, which is why some of them achieve even greater success than those born into wealthy families.

These comments indicate that some students felt that people can succeed despite growing up in a disadvantaged environment. They emphasized both the difficulty of achieving success and the importance of personal effort. Overall, these responses reveal an inherent conflict between acknowledging structural inequalities and maintaining a belief in the importance of individual effort and upward mobility. Some students emphasized the disadvantages associated with poverty, while others stressed the possibility of overcoming such barriers through determination, education, and hard work.

7.4 INTERNATIONAL STUDENTS' CRITICAL RESPONSES (DISAGREEMENT WITH "OYA GACHA")

The co-occurrence network analysis categorized international students' critical views on "Oya Gacha" into three themes (Figure 4).

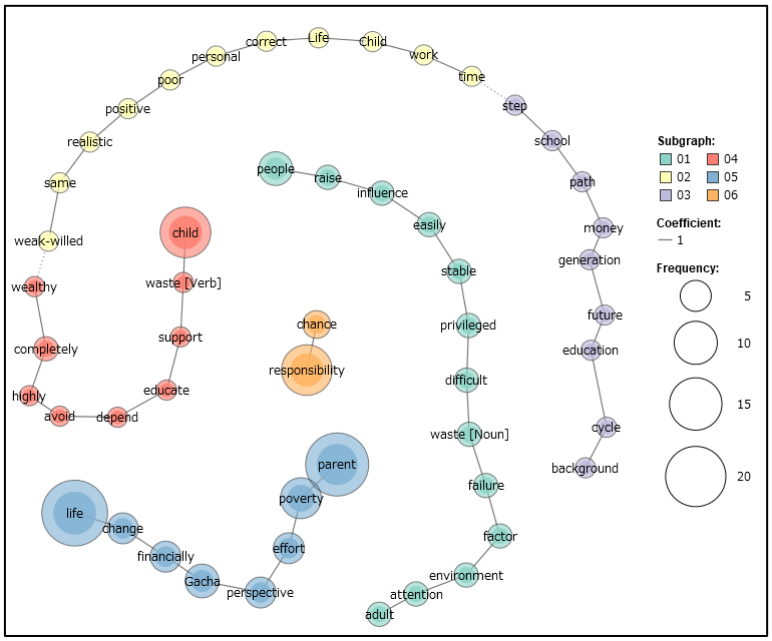


Figure 4: Co-occurrence Network of International Students' Critical Responses

7.4.1 CONCERNS ABOUT BEING TOO DEPENDENT ON PARENTS

Some students argued that regardless of parental wealth, children should not rely entirely on their parents for success, claiming that hard work can change one's circumstances. One international student stated the following:

A child who is born into a financially poor family can change their future through education and effort. A child who is born into a wealthy family may waste opportunities, avoid work, and fall into poverty.

This comment is an example of the opinion held by some that parental financial status does not determine one's future. Another international student emphasized the importance of self-reliance: "Even with bad parents, effort can change one's life. Adulthood means taking responsibility. Blaming parents is a waste." This comment denotes the importance that some students place on the value of self-determination. These responses suggest that some students view the concept of "Oya Gacha" as problematic because it could lead to excessive reliance on one's family background. Instead, they emphasize the importance of individual responsibility, effort, and self-reliance in determining life outcomes.

7.4.2 COMBINED INFLUENCE OF ENVIRONMENTAL FACTORS AND PERSONAL RESPONSIBILITY

Some students highlighted that one's environment and personal choices both shape success. Effort can create opportunities, even for those experiencing poverty. However, some see this view as being inherently privileged and the product of a stable upbringing. One international student stated, "People can't choose their parents, but is it fair to blame them? Life depends on many factors, and effort can change one's circumstances." This comment conveys that life is shaped not only by parents but also by external and individual factors, such as access to education and personal choices.

Another student noted, "Poverty is cyclical. Some parents blame their children, saying 'Why are you so weak-willed?'" This comment highlights the intergenerational nature of poverty, indicating that success is about more than effort, while recognizing opportunities for self-improvement are often available. These responses also suggest that students acknowledge structural constraints and individual agency as significant factors that influence life outcomes. According to this view, success is shaped by the interaction between social circumstances and personal effort rather than by either factor alone.

7.4.3 OVERCOMING PARENTAL INFLUENCE

Some students argued that parental influence is real, but not an excuse to give up. The "Oya Gacha" mindset may discourage ambition and overlook the value of effort. Students emphasized the principle that, even in difficult economic conditions, people can change their lives through effort. As one student commented, "Parents and children shouldn't rely on others. Each person should live independently, and mutual support is key to escaping poverty." This comment conveys the value of self-reliance, suggesting that the "Oya Gacha" mindset limits effort and constrains life choices, whereas escaping poverty requires independent action. These responses suggest that some students view "Oya Gacha" as pessimistic and an underestimation of the role of individual agency. Instead, they emphasize the importance of independence, perseverance, and personal responsibility in overcoming disadvantaged circumstances.

8. DISCUSSION AND CONCLUSION

This study examined the perceptions of Japanese and international students regarding "Oya Gacha" through the lens of cultural reproduction theory. Analyses of quantitative and qualitative data revealed common perceptions and significant differences that appeared to be shaped by the cultural and institutional contexts in which the participants were socialized.

8.1 CULTURAL REPRODUCTION AND AGREEMENT WITH "OYA GACHA"

This study found strong agreement with the concept of "Oya Gacha" among both groups. Quantitative analysis revealed that Japanese ($M = 3.16$) and international ($M = 3.23$) students exhibited similar mean levels of agreement, and the difference between the two groups was not significant. These results suggest that the recognition of how family background influences life outcomes is widely shared among students, regardless of their national background. In Japan, access to private and cram schools depends significantly on parental wealth and often determines children's educational opportunities. This is consistent with cultural reproduction theory, which links family wealth and education to academic achievement (Sengonul, 2022). Students from low-income families often lack the social and academic knowledge necessary to succeed, thereby widening the educational gap between rich and poor (White et al., 2020). Qualitative analysis showed that many students recognized the impact of parental wealth on college admissions and careers. In particular, Japan's private and cram school culture appears to reinforce economic and educational inequalities. This supports Sengonul's (2022) argument that family wealth promotes academic success. The small effect size observed in the quantitative analysis indicates that the level of agreement with the concept of "Oya Gacha" is similar across both groups despite their different cultural and national contexts.

Japanese students' views were more consistent than those of international students. Although the mean scores were similar, the higher SD and CV values among international students suggest that their views were more diverse and polarized. This suggests that cultural reproduction is widely accepted in Japan; however, using data on student backgrounds may help discern whether perceptions vary because of societal and educational differences. Some international students agreed strongly with "Oya Gacha," whereas others did not. This pattern suggests that although many students are familiar with the concept of "Oya Gacha," their interpretations may differ based on their national, cultural, and educational backgrounds. These quantitative results provide important statistical context for understanding the qualitative responses presented below.

Broadly, these findings align with international research on the intergenerational reproduction of inequality. Previous studies have demonstrated that family resources influence access to educational opportunities and

students' confidence, aspirations, and ability to navigate institutional expectations. From this perspective, the concept of "Oya Gacha" can be viewed as a local expression of a global pattern: the unequal transmission of economic, cultural, and social capital across generations. Thus, this study contributes to ongoing international discussions on social inequality by illustrating how Japanese and international students articulate these concerns through culturally specific terms and narratives.

8.2 AWARENESS OF SOCIAL INEQUALITY AND INTERNATIONAL DIFFERENCES

The qualitative analysis revealed that Korean students referred to the concept of a "gold spoon." Similar to the Japanese students, they recognized that parental finances significantly influence educational and career opportunities. This is consistent with research showing that differences in parental involvement in education between families from different social backgrounds contribute to educational inequality (Kosutic, 2017; Sengonul, 2022). However, some international students believed that individual effort could overcome parental influence.

Compared to Japanese students, a higher percentage of international students (43.3%) strongly agreed with the concept of "Oya Gacha." This suggests that international students may be more affected by social inequality in their home countries. In contrast, Japanese students were more aware of how social structures reinforce "Oya Gacha." This difference is significant in the context of comparative research on social inequality, which indicates that attitudes toward meritocracy, mobility, and personal responsibility differ across national contexts. In societies where upward mobility is perceived as achievable, students may emphasize individual effort and personal initiative. Conversely, in countries where educational competition and class-based inequalities are highly visible, students may consider family background to be the decisive factor in one's chances in life.

8.3 INDIVIDUAL EFFORT AND THE EFFECTS OF SOCIAL SYSTEMS

Many students criticized the concept of "Oya Gacha." Both groups included students who claimed that effort shapes one's future and parental influence is exaggerated. This may be because, for those from advantaged backgrounds, the education system appears to be fair and neutral, with such individuals failing to see how it reflects the existing power structure in society and gives advantages to children who are more familiar with the dominant culture (Kosutic, 2017). Many Japanese students acknowledged that although the family environment is important, life outcomes can be affected by individual effort. This suggests that they understood cultural reproduction theory but emphasized personal responsibility as much or more than parental influence. Some students argued that the core problem is not parents but the social system itself and advocated for broader social reforms.

Similarly, international students who rejected the concept of "Oya Gacha" believed that while environmental factors play a role, they do not determine success. Students from high-mobility societies emphasized effort over parental influence in achieving success. They also noted that scholarship programs and social support systems are crucial factors in overcoming the effects of parental financial status (Sengonul, 2022). Securing funding for college is a critical factor in academic success and a significant source of stress (White et al., 2020). Therefore, expanding scholarship programs and widening access to social support systems are necessary to improve access to educational opportunities for students from economically disadvantaged backgrounds.

The findings are also related to a broader debate in global inequality research on the relationship between structural constraints and individual effort. Several respondents acknowledged the existence of inequality while questioning whether the term "Oya Gacha" might diminish people's sense of personal responsibility. This ambivalence reflects international findings showing that students from disadvantaged backgrounds often experience structural barriers and strong expectations to succeed through merit and effort. Therefore, this study suggests that "Oya Gacha" should not be understood as merely a pessimistic belief and may instead function as a social expression through which students reflect on the relationship between family background and individual effort.

8.4 FUTURE RESEARCH

Future studies should include gender and country data to analyze how these factors shape perceptions of "Oya Gacha." Additionally, future research should collect a wide range of demographic and socioeconomic variables, such as nationality, gender, parental education level, household income, first-generation college student status, scholarship status, type of high school attended, and urban–rural background. Incorporating these variables would enable precise examination of how students' social positions shape their interpretations of educational inequality and whether perceptions of the concept of "Oya Gacha" vary systematically across social and demographic groups.

9. IMPLICATIONS

This study examined perceptions of educational inequalities in Japan and the validity of the concept of "Oya Gacha." It analyzed attitudinal differences between Japanese and international students using cultural

reproduction theory. The implications of this study are summarized below.

9.1 JAPANESE SOCIAL ISSUES EXPLAINED BY CULTURAL AND SOCIAL REPRODUCTION THEORY

This study found that many Japanese students believe that family wealth strongly influences educational and career opportunities, thus perpetuating social inequality. Okabe (2025) noted that SES at birth affects academic achievement and access to education, warning that overemphasizing personal effort risks misleading future leaders. He emphasized the need for future leaders to understand educational inequality. Japanese students strongly linked parental wealth to academic and career success, confirming the popularity and salience of cultural reproduction theory as a framework for understanding Japanese society. Furthermore, including international students' views provided a global perspective on educational inequality in Japan. Some international students believed that parental influence could be overcome through effort, while others disagreed, which may depend on the equality levels of their own countries.

9.2 SOCIAL SIGNIFICANCE OF "OYA GACHA" AND POSSIBILITIES FOR CHANGING AWARENESS

This study demonstrated that "Oya Gacha" refers to more than personal dissatisfaction; it highlights social inequality. For many Japanese students, "Oya Gacha" also invites structural reforms. In Japan, where personal responsibility is emphasized, "Oya Gacha" encourages reflection on structural constraints to success. International students' views varied by country. The results suggested that those from low-mobility societies consider parental wealth to be the decisive factor, whereas those from high-mobility societies value individual effort more. This suggests that views on "Oya Gacha" vary with differences in social structures. This study can raise awareness of the role of parental influence and social structures in fostering social injustice, and future studies should examine different national and social backgrounds to assess how nationality and social context affect perceptions of the significance of parental wealth in determining life outcomes. At the international level, the concept of "Oya Gacha" could be a useful starting point for comparative research on educational inequality. Similar ideas appear in other societies under different terms, such as inherited advantage, family background effects, and class reproduction. Examining how these concepts are discussed in different countries could help researchers better understand how students perceive fairness, social mobility, and social inequality in various cultural contexts. Future studies should incorporate multivariate analyses using more extensive demographic data to determine if agreement with the concept of "Oya Gacha" is more strongly associated with nationality, family SES, parental education, or first-generation status.

REFERENCES

Bourdieu, P., & Passeron, J. C. (1977). *Reproduction in education, society and culture*. Sage Publications.

Grant, P. D., & Kniess, D. (2023). 'I just kind of felt like country come to town:' College student experiences for rural students at one flagship university. *The Rural Educator*, 44(3), 1–16. https://doi.org/10.55533/2643-9662.1357

Higuchi, K. (2016). A two-step approach to quantitative content analysis: KH Coder tutorial using Anne of Green Gables (Part I). *Ritsumeikan Social Science Review* 52(3), 77–91. http://www.ritsumei.ac.jp/file.jsp?id=325881

Higuchi, K. (2017). A two-step approach to quantitative content analysis: KH Coder tutorial using Anne of Green Gables (Part II). *Ritsumeikan Social Science Review*, 53(1), 137–147. https://www.ritsumei.ac.jp/file.jsp?id=346128

Hida, M. (2021). "The concept of Oya Gacha is correct" – The reason an American economist does not deny the lottery of life. ["Oya Gacha toiu gainen wa tadashii." Americajin keizaigakusha ga jinsei no takarakuji wo hiteishinai wake] PRESIDENT Online. December 5, 2021. Accessed December 20, 2024. https://president.jp/articles/-/52516?page=1 [Published in Japanese]

Jiyu Kokuminsha. (2021). Basic Knowledge of the Year Book of the Contemporary Society Selection - U-CAN New Words & Buzzwords Award - 38th Edition - 2021 Winning Words. Accessed January 12, 2024. https://www.jiyu.co.jp/singo/index.php?eid=00038 [published in Japanese]

Kosutic, I. (2017). The role of cultural capital in higher education access and institutional choice. *CEPS Journal*, 7(1),149–169.

Okabe, T. (2025). "Which high school you go to" is actually determined by your birth. Examining educational inequality through the unique 'high school realities' of Kyoto University Students." ["Dokono Koukou ni ikuka" wa, jitsuwa "Umare" ga kimeteiru.Tokushu sugiru Kyoodaisei "Koukou aruaru" kara kyouiku kakusa wo kanngaeta.]. JBpress. January 21, 2025. Accessed January 25, 2025. https://jbpress.ismedia.jp/articles/-/84282 [Published in Japanese]

Sengonul, T. (2022). A review of the relationship between parental involvement and children's academic achievement and the role of family socioeconomic status in this relationship. *Pegem Journal of Education*

and Instruction, 12(2), 32–57.
White, J. W., Pascale, A., & Aragon, S. (2020). Collegiate cultural capital and integration into the college community. *College Student Affairs Journal*, 38(1), 34–52.

Teachers' Perceptions of Generative AI in Education: Opportunities, Challenges, and Classroom Use

Ali AKÇİN
İskele Evkaf Türk Maarif Koleji
aliakcin82@gmail.com

Kemal KURUKAFA
İskele Evkaf Türk Maarif Koleji
kemal.kurukafa@hotmail.com

Adil TARHAN
Dipkarpaz Recep Tayyip Erdoğan Ortaokulu
tarhan01adil@gmail.com

ABSTRACT
This qualitative study explores teachers' perceptions of generative artificial intelligence (AI) in education, focusing on opportunities, challenges, and classroom applications. Using semi-structured interviews with 45 teachers from primary, secondary, and high school levels across Turkey, the research employed thematic analysis to identify key themes regarding generative AI integration. The findings suggest that teachers recognize significant opportunities including personalized learning, enhanced content creation, time efficiency, and improved student engagement. That said, they also express substantial concerns about academic integrity, over-reliance on technology, inadequate training, ethical implications, and equity issues. Teachers report diverse classroom applications ranging from lesson planning and assessment creation to providing feedback and generating instructional materials. This research identifies six major themes: (1) Perceived Opportunities and Benefits, (2) Challenges and Barriers to Implementation, (3) Current Classroom Applications, (4) Professional Development Needs, (5) Ethical and Academic Integrity Concerns, and (6) Future Integration Perspectives. Results indicate that successful generative AI integration requires comprehensive professional development, clear institutional policies, and balanced approaches that preserve critical thinking while leveraging technological advantages. This research contributes to understanding how educators perceive and utilize generative AI tools, offering insights for policymakers, educational institutions, and teacher training programs.
Keywords: Generative artificial intelligence, teachers' perceptions, educational technology, classroom integration, ChatGPT, opportunities and challenges

1. INTRODUCTION

The emergence of generative artificial intelligence (AI) technologies, particularly large language models such as ChatGPT, has created unprecedented opportunities and challenges in educational contexts worldwide (Kasneci et al., 2023). These AI systems, capable of generating human-like text, creating educational content, and providing instant feedback, have rapidly transformed teaching and learning practices across all educational levels (Baidoo-Anu & Owusu Ansah, 2023). As generative AI tools become increasingly accessible and sophisticated, understanding teachers' perceptions of these technologies has become critical for effective educational integration and policy development (Cotton et al., 2023).

Generative AI represents a paradigm shift in educational technology, moving beyond traditional computer-assisted instruction to systems that can engage in complex reasoning, generate original content, and adapt to individual learner needs (Mollick & Mollick, 2023). Unlike previous educational technologies, generative AI tools offer unprecedented capabilities for personalization, content creation, and instructional support (Rudolph et al., 2023). That said, their rapid adoption has also raised significant concerns about academic integrity, critical thinking development, and the changing role of educators in AI-enhanced learning environments (Zawacki-Richter et al., 2019).

Teachers stand at the forefront of this technological transformation, serving as key decision-makers in determining how, when, and whether to integrate generative AI into their pedagogical practices (Crompton & Burke, 2023). Their perceptions, attitudes, and experiences significantly influence the success or failure of technology integration initiatives (Lim et al., 2023). Research indicates that teachers' beliefs about technology directly impact their adoption behaviors, instructional strategies, and ultimately, student learning outcomes

(Celik, 2023). As a result, examining teachers' perceptions of generative AI provides essential insights into both the opportunities and obstacles facing educational institutions as they navigate this technological revolution.

Despite growing interest in generative AI in education, significant gaps remain in our understanding of teachers' perspectives, particularly regarding practical classroom applications and the specific challenges educators face when implementing these technologies (Ellis et al., 2025). While several studies have examined student use of AI tools or focused on higher education contexts, fewer have explored K-12 teachers' comprehensive perceptions across different school levels (Kim, 2025). In addition, most existing research has been conducted in Western contexts, with limited attention to teachers' experiences in diverse cultural and educational settings (Majeed et al., 2024).

In this study addresses these gaps by investigating teachers' perceptions of generative AI across primary, secondary, and high school levels, with particular attention to the opportunities they identify, the challenges they encounter, and the ways they currently use or envision using these tools in their classrooms. The research is guided by three primary questions: (1) What opportunities and benefits do teachers perceive in generative AI for education? (2) What challenges and concerns do teachers identify regarding generative AI integration? (3) How are teachers currently using or planning to use generative AI in their classroom practices? By examining these questions through in-depth qualitative inquiry, this study aims to provide a nuanced understanding of teachers' perspectives that can inform professional development programs, institutional policies, and future research on AI in education (Tongchai et al., 2024).

2. METHODOLOGY

2.1. Research Design

In this study employed a qualitative research design using semi-structured interviews to explore teachers' perceptions of generative AI in education. The qualitative approach was selected because it allows for in-depth exploration of participants' experiences, beliefs, and attitudes, providing rich contextual understanding that quantitative methods alone cannot capture (Creswell & Poth, 2018). Semi-structured interviews were chosen as the primary data collection method because they offer flexibility to probe interesting responses while maintaining consistency across interviews through a predetermined set of core questions (Merriam & Tisdell, 2016).

The research was grounded in interpretive phenomenology, which seeks to understand how individuals make sense of their lived experiences (Van Manen, 2016). This philosophical approach is particularly appropriate for investigating teachers' perceptions because it acknowledges that understanding of technology is socially constructed and context-dependent (Patton, 2015). This research aimed to capture the complexity and diversity of teachers' perspectives rather than seeking to generalize findings to all educational contexts.

2.2. Study Group and Sampling

This research group consisted of 45 teachers from various educational levels in Turkey, selected through purposive sampling. Purposive sampling was employed to ensure representation across different school levels, subject areas, and levels of experience with generative AI technologies (Palinkas et al., 2015). The sample included 15 primary school teachers (grades 1-4), 15 secondary school teachers (grades 5-8), and 15 high school teachers (grades 9-12) from both public and private schools in Ankara and surrounding provinces.

Participant selection criteria included: (1) at least three years of teaching experience, (2) awareness of generative AI tools such as ChatGPT, and (3) willingness to discuss their perceptions and experiences openly. The sample included teachers from diverse subject areas including Turkish language and literature, mathematics, science, social studies, English language, and information technology. Among the 45 participants, 28 were female and 17 were male, with teaching experience ranging from 3 to 24 years (mean = 11.3 years). Approximately 60% of participants reported having used generative AI tools at least once for professional purposes, while 40% had not yet used these tools but were aware of their existence and capabilities.

2.3. Data Collection Tool

Data were collected through semi-structured interviews conducted between September 2025 and January 2026. The interview protocol was developed based on extensive literature review and consultation with experts in educational technology and qualitative research methods (Rasool et al., 2025; Wang, 2024). The protocol consisted of 18 open-ended questions organized into five main sections: (1) awareness and familiarity with generative AI, (2) perceived opportunities and benefits, (3) challenges and concerns, (4) current and potential classroom applications, and (5) professional development needs and future perspectives.

Sample interview questions included: "What is your understanding of generative AI tools like ChatGPT?", "What opportunities do you see for using generative AI in your teaching practice?", "What concerns do you have about integrating generative AI in education?", "How have you used or how might you use generative AI in your classroom?", and "What support would you need to effectively integrate generative AI into your teaching?" The interview protocol was piloted with three teachers not included in the final sample, and minor revisions were made based on their feedback to improve clarity and flow.

2.4. Data Collection Process

Interviews were conducted individually, either face-to-face in participants' schools or via video conferencing platforms based on participant preference and logistical considerations. Each interview lasted between 45 and 75 minutes (average = 58 minutes). All interviews were audio-recorded with participants' explicit consent and transcribed verbatim within 48 hours of completion. Field notes were also taken during and immediately after each interview to capture non-verbal cues, contextual information, and initial impressions.

Prior to data collection, ethical approval was obtained from the Ankara University Ethics Committee (Approval No: 2025-08-142). Participants were provided with detailed information about the study's purpose, procedures, and their rights, and all provided written informed consent. Confidentiality was ensured through the use of pseudonyms and removal of identifying information from transcripts. Participants were informed of their right to withdraw from the study at any time without consequences.

2.5. Data Analysis

Data analysis followed Braun and Clarke's (2006) six-phase thematic analysis approach: (1) familiarization with the data through repeated reading of transcripts, (2) generation of initial codes, (3) searching for themes by grouping codes, (4) reviewing and refining themes, (5) defining and naming themes, and (6) producing the final analysis. The analysis was conducted using NVivo 14 qualitative data analysis software to facilitate systematic coding and theme development.

The coding process began with open coding, where segments of text were assigned descriptive codes capturing their essential meaning. Initial coding resulted in 127 distinct codes. These codes were then grouped into broader categories based on conceptual similarities, resulting in 18 categories. Through iterative review and refinement, these categories were organized into six major themes and 14 sub-themes. Two researchers independently coded 20% of the transcripts to establish inter-coder reliability, achieving a Cohen's kappa coefficient of 0.84, indicating strong agreement (McHugh, 2012).

2.6. Validity and Reliability

Multiple strategies were employed to enhance the trustworthiness of the research findings (Lincoln & Guba, 1985). Credibility was established through prolonged engagement with the data, triangulation of perspectives across different school levels and subject areas, and member checking with 12 participants who reviewed their interview transcripts and preliminary findings. Transferability was addressed through thick description of the research context, participants, and findings, allowing readers to assess the applicability of results to their own contexts.

Dependability was ensured through maintaining a detailed audit trail documenting all research decisions, coding processes, and analytical steps. An external auditor with expertise in qualitative research reviewed the audit trail and confirmed the logical consistency of the research process. Confirmability was enhanced through reflexive journaling, where researchers documented their assumptions, biases, and reactions throughout the research process, and through the use of direct quotations from participants to ground findings in the data. These measures collectively strengthen confidence in the research findings and their interpretation.

3. FINDINGS

The thematic analysis of interview data revealed six major themes regarding teachers' perceptions of generative AI in education: (1) Perceived Opportunities and Benefits, (2) Challenges and Barriers to Implementation, (3) Current Classroom Applications, (4) Professional Development Needs, (5) Ethical and Academic Integrity Concerns, and (6) Future Integration Perspectives. Each theme is presented below with supporting data and illustrative quotations from participants.

3.1. Theme 1: Perceived Opportunities and Benefits

Teachers identified numerous opportunities and benefits associated with generative AI in education. Table 1 presents the frequency and percentage of teachers mentioning specific opportunities.

Table 1. Perceived Opportunities and Benefits of Generative AI (N=45)

Opportunity/Benefit	Frequency	Percentage
Personalized learning and differentiation	38	84.4%
Time-saving in lesson planning and preparation	35	77.8%
Enhanced content creation and material development	33	73.3%
Immediate feedback provision	29	64.4%
Increased student engagement and motivation	27	60.0%
Support for diverse learning needs	25	55.6%
Access to diverse resources and examples	23	51.1%
Language learning support	19	42.2%
Administrative task automation	17	37.8%

The most frequently mentioned opportunity was personalized learning and differentiation (84.4%), with teachers recognizing generative AI's potential to tailor content and activities to individual student needs (Javahery et al., 2025). One high school mathematics teacher explained: "With AI, I can generate different versions of the same problem at varying difficulty levels. This allows me to challenge advanced students while providing appropriate support for those who struggle." Time-saving benefits in lesson planning and preparation were mentioned by 77.8% of participants, reflecting teachers' appreciation for AI's efficiency in generating ideas, materials, and resources (Reis-Andersson, 2024). A primary school teacher noted: "ChatGPT helps me brainstorm creative activities and generate examples much faster than searching through textbooks or websites."

Enhanced content creation capabilities were identified by 73.3% of teachers, who valued AI's ability to generate explanations, examples, worksheets, and assessment items (Escario et al., 2024). Teachers also recognized opportunities for providing immediate feedback (64.4%) and increasing student engagement through interactive and varied learning experiences (60.0%). Several teachers mentioned AI's potential to support students with diverse learning needs, including English language learners and students with learning disabilities, by providing scaffolded explanations and multiple representations of concepts (Saud, 2025).

3.2. Theme 2: Challenges and Barriers to Implementation

Despite recognizing opportunities, teachers identified substantial challenges and barriers to implementing generative AI in their classrooms. Table 2 summarizes the main challenges mentioned by participants.

Table 2. Challenges and Barriers to Generative AI Implementation (N=45)

Challenge/Barrier	Frequency	Percentage
Academic integrity and plagiarism concerns	41	91.1%
Student over-reliance and reduced critical thinking	37	82.2%
Inadequate teacher training and preparation	34	75.6%
Accuracy and reliability of AI-generated content	31	68.9%
Lack of institutional policies and guidelines	28	62.2%
Digital divide and equity issues	26	57.8%
Ethical concerns and bias in AI outputs	24	53.3%
Difficulty in assessment and evaluation	22	48.9%
Technological infrastructure limitations	18	40.0%

Academic integrity and plagiarism concerns were mentioned by 91.1% of teachers, representing the most significant barrier to AI integration (Pettersson et al., 2024). A secondary school Turkish language teacher expressed: "My biggest worry is that students will simply copy AI-generated essays without learning anything. How can I assess their actual writing skills?" Teachers' concerns about student over-reliance and reduced critical thinking were also prevalent (82.2%), with many fearing that easy access to AI-generated answers might discourage deep thinking and problem-solving (Munawar et al., 2024).

Inadequate teacher training emerged as a critical barrier (75.6%), with most teachers reporting they had received no formal professional development on generative AI (Moorhouse et al., 2024). One experienced teacher stated: "I know these tools exist, but I don't really understand how they work or how to use them effectively in my teaching. We need proper training." Concerns about the accuracy and reliability of AI-generated content were mentioned by 68.9% of participants, who worried about potential misinformation or errors in AI outputs (Kim et al., 2022). The lack of clear institutional policies and guidelines (62.2%) left many teachers uncertain about appropriate uses and boundaries for AI in their schools (Soleimani et al., 2025).

3.3. Theme 3: Current Classroom Applications

Teachers reported diverse ways they were currently using or envisioning using generative AI in their classroom practices. Table 3 presents the frequency of different classroom applications.

Table 3. Current and Potential Classroom Applications of Generative AI (N=45)

Classroom Application	Frequency	Percentage
Lesson planning and instructional design	32	71.1%
Creating practice questions and exercises	28	62.2%
Generating examples and explanations	27	60.0%
Developing assessment items and rubrics	24	53.3%
Providing writing feedback and suggestions	21	46.7%
Creating differentiated materials	19	42.2%
Translating and simplifying content	17	37.8%
Brainstorming and idea generation	16	35.6%
Creating visual aids and presentations	14	31.1%

Lesson planning and instructional design were the most common applications (71.1%), with teachers using AI to generate lesson ideas, structure activities, and develop teaching materials (Nyaaba, 2024). A middle school science teacher explained: "I use ChatGPT to help me plan inquiry-based lessons. It suggests interesting questions and activities that I can adapt for my students." Creating practice questions and exercises was mentioned by 62.2% of teachers, who valued AI's ability to quickly generate varied practice items at different difficulty levels (Reis-Andersson, 2024).

Teachers also used generative AI to generate multiple explanations of complex concepts (60.0%), develop assessment items and rubrics (53.3%), and provide feedback on student writing (46.7%). Several teachers mentioned using AI to create differentiated materials for students with varying abilities (42.2%) and to translate or simplify content for English language learners (37.8%). A high school English teacher noted: "I use AI to generate different versions of reading passages at various reading levels, which helps me support all my students" (Ruediger et al., 2024).

3.4. Theme 4: Professional Development Needs

Teachers expressed clear needs for professional development to effectively integrate generative AI into their teaching practices. Table 4 summarizes the specific professional development needs identified by participants.

Table 4. Professional Development Needs for Generative AI Integration (N=45)

Professional Development Need	Frequency	Percentage
Technical skills for using AI tools effectively	39	86.7%
Pedagogical strategies for AI integration	36	80.0%
Evaluating AI-generated content for accuracy	33	73.3%
Addressing ethical issues and academic integrity	31	68.9%
Designing AI-enhanced assessments	27	60.0%
Understanding AI capabilities and limitations	25	55.6%
Promoting critical thinking alongside AI use	23	51.1%
Equity and access considerations	19	42.2%

The most frequently mentioned need was for technical skills training (86.7%), with teachers wanting hands-on experience with various AI tools and practical guidance on their features and functions (Wardat et al., 2025). That said, teachers emphasized that technical training alone was insufficient; they also needed pedagogical strategies for meaningful AI integration (80.0%). A primary school teacher stated: "I can learn to use the technology, but I need help understanding when and how to use it in ways that actually improve student learning" (Echave et al., 2024).

Teachers also expressed strong needs for training on evaluating AI-generated content for accuracy and appropriateness (73.3%), addressing ethical issues and academic integrity concerns (68.9%), and designing assessments that remain valid in an AI-enhanced environment (60.0%). Several teachers mentioned wanting to better understand AI's capabilities and limitations (55.6%) to set realistic expectations and use tools appropriately. One teacher emphasized: "We need ongoing professional development, not just a one-time workshop. This technology is evolving so quickly" (Yue et al., 2025).

3.5. Theme 5: Ethical and Academic Integrity Concerns

Ethical considerations and academic integrity concerns emerged as a major theme, with teachers expressing complex and sometimes conflicting views about AI's impact on learning and assessment. Table 5 presents the specific ethical concerns mentioned by participants.

Table 5. Ethical and Academic Integrity Concerns (N=45)

Ethical Concern	Frequency	Percentage
Student plagiarism and cheating	40	88.9%
Difficulty detecting AI-generated work	35	77.8%
Undermining authentic learning	32	71.1%
Bias and fairness in AI outputs	28	62.2%
Privacy and data security	24	53.3%
Transparency about AI use	22	48.9%
Devaluing human creativity and thinking	20	44.4%
Unequal access creating advantages	18	40.0%

Student plagiarism and cheating were the most prominent ethical concerns (88.9%), with teachers worried that AI makes it too easy for students to submit work that is not their own (Sagocsoc et al., 2025). A high school history teacher expressed: "How can I know if an essay was written by the student or by ChatGPT? Traditional plagiarism detection doesn't work anymore." The difficulty of detecting AI-generated work (77.8%) compounded these concerns, leaving teachers feeling uncertain about the authenticity of student submissions (Wang, 2024).

Many teachers worried that AI use might undermine authentic learning (71.1%), with students receiving answers without engaging in the cognitive processes necessary for deep understanding. Concerns about bias and fairness in AI outputs were mentioned by 62.2% of participants, who worried about perpetuating stereotypes or providing culturally inappropriate content (Pettersson et al., 2024). Privacy and data security concerns (53.3%) were also raised, particularly regarding student information being processed by commercial AI systems. Teachers emphasized the need for transparency about AI use (48.9%), with clear expectations about when and how students should disclose AI assistance (Ututalum, 2025).

3.6. Theme 6: Future Integration Perspectives

Teachers expressed diverse perspectives on the future role of generative AI in education, ranging from enthusiastic optimism to cautious skepticism. Table 6 summarizes teachers' views on future AI integration.

Table 6. Perspectives on Future Generative AI Integration (N=45)

Future Perspective	Frequency	Percentage
AI will become essential teaching tool	29	64.4%
Need for balanced human-AI collaboration	38	84.4%
Curriculum and assessment must adapt	33	73.3%
Teacher role will evolve, not disappear	35	77.8%
Institutional policies are critical	31	68.9%
Student AI literacy is essential	27	60.0%
Cautious, gradual implementation preferred	25	55.6%
Concerns about widening achievement gaps	21	46.7%

While 64.4% of teachers believed generative AI would become an essential teaching tool, the overwhelming majority (84.4%) emphasized the need for balanced human-AI collaboration rather than AI replacement of human teaching (Ellis et al., 2025). A middle school teacher articulated this view: "AI should be a tool that enhances what we do, not something that replaces the human connection and judgment that are essential to good teaching." Many teachers (73.3%) recognized that curriculum and assessment practices would need to adapt to AI's presence, with traditional assignments potentially becoming obsolete (Kim, 2025).

Teachers generally believed their role would evolve rather than disappear (77.8%), with increased emphasis on facilitating critical thinking, creativity, and ethical reasoning that AI cannot replicate. The importance of institutional policies was stressed by 68.9% of participants, who wanted clear guidelines about appropriate AI use for both teachers and students (Majeed et al., 2024). Teachers also emphasized the need to develop students' AI literacy (60.0%), teaching them to use these tools responsibly and critically. Many teachers (55.6%) preferred cautious, gradual implementation with ongoing evaluation rather than rapid, wholesale adoption. Some

teachers (46.7%) expressed concerns that AI might widen achievement gaps if access and support were not equitably distributed (Chung et al., 2024).

4. DISCUSSION

The findings of this study reveal a complex landscape of teachers' perceptions regarding generative AI in education, characterized by recognition of significant opportunities alongside substantial concerns and uncertainties. This discussion situates the findings within existing literature and explores their implications for educational practice and policy.

Teachers' strong recognition of personalized learning opportunities aligns with previous research highlighting AI's potential for differentiation and individualized instruction (Mollick & Mollick, 2023). The high percentage of teachers (84.4%) identifying personalization as a key benefit suggests that educators understand AI's capacity to address one of education's most persistent challenges. That said, the gap between perceived potential and actual implementation indicates that realizing this potential requires more than technological availability. This finding echoes Soleimani et al. (2025), who found that positive attitudes toward AI do not automatically translate into effective classroom integration without adequate support and pedagogical guidance.

The prominence of academic integrity concerns (91.1%) reflects a central tension in current educational discourse about generative AI. Teachers' worries about plagiarism and cheating are well-founded, as research has documented students' use of AI to complete assignments without genuine learning (Cotton et al., 2023). That said, these concerns also reflect a need to reconceptualize assessment practices in an AI-enhanced environment. As Rudolph et al. (2023) argue, rather than attempting to prevent AI use through detection and prohibition, educators might better serve students by redesigning assessments to emphasize higher-order thinking, creativity, and authentic application that AI cannot easily replicate.

The widespread lack of adequate teacher training (75.6%) represents a critical barrier to effective AI integration and aligns with findings from multiple international studies (Moorhouse et al., 2024; Wardat et al., 2025). Teachers' emphasis on needing both technical skills and pedagogical strategies reflects a sophisticated understanding that technology integration requires more than operational competence. This finding supports the Technological Pedagogical Content Knowledge (TPACK) framework, which emphasizes the intersection of technology, pedagogy, and content knowledge in effective teaching (Mishra & Koehler, 2006). Professional development programs must address all three domains, helping teachers understand not just how to use AI tools, but when, why, and in what ways they enhance specific learning objectives.

Teachers' emphasis on balanced human-AI collaboration (84.4%) rather than AI replacement reflects a mature understanding of technology's role in education. This perspective aligns with research emphasizing that effective educational technology augments rather than replaces human teaching (Zawacki-Richter et al., 2019). The finding that 77.8% of teachers believe their role will evolve rather than disappear suggests resilience and adaptability in the teaching profession. The ethical concerns teachers raised, particularly regarding bias, privacy, and equity, reflect growing awareness of AI's societal implications (Selwyn, 2019). The finding that 57.8% of teachers identified digital divide and equity issues as barriers underscores the need for implementation strategies that prioritize equitable access and culturally responsive AI use.

5. RESULTS AND CONCLUSIONS

This qualitative study investigated teachers' perceptions of generative AI in education through semi-structured interviews with 45 teachers across primary, secondary, and high school levels. The research identified six major themes: Perceived Opportunities and Benefits, Challenges and Barriers to Implementation, Current Classroom Applications, Professional Development Needs, Ethical and Academic Integrity Concerns, and Future Integration Perspectives.

Teachers identified numerous opportunities associated with generative AI, most prominently personalized learning and differentiation (84.4%), time-saving in lesson planning (77.8%), and enhanced content creation (73.3%). That said, teachers also expressed significant concerns, particularly regarding academic integrity and plagiarism (91.1%), student over-reliance and reduced critical thinking (82.2%), and inadequate teacher training (75.6%). Current classroom applications focus primarily on teacher tasks such as lesson planning (71.1%), creating practice questions (62.2%), and generating examples and explanations (60.0%), while student-centered AI applications remain underutilized. Teachers expressed clear professional development needs, particularly for technical skills (86.7%), pedagogical strategies (80.0%), and guidance on evaluating AI-generated content (73.3%).

Several important conclusions emerge from this research. First, successful generative AI integration requires more than technological access; it demands comprehensive professional development, clear institutional policies, and pedagogical innovation. Second, teachers' perspectives are diverse and context-dependent, suggesting that flexible frameworks allowing local adaptation are preferable to uniform implementation approaches. Third, teachers are thoughtful, critical consumers of educational technology whose concerns about academic integrity, critical thinking, and equity should inform implementation strategies rather than be dismissed as resistance to change. Finally, the emphasis on balanced human-AI collaboration suggests that the future of education lies in teachers effectively leveraging AI as one powerful tool within their broader pedagogical repertoire, rather than AI replacing the irreplaceable human dimensions of teaching and learning.

6. RECOMMENDATIONS

6.1. Practical Recommendations

1. Develop Comprehensive Professional Development Programs: Educational institutions should create multi-tiered, ongoing professional development programs addressing both technical skills and pedagogical strategies for generative AI integration, providing teachers with opportunities for hands-on practice and peer collaboration.
2. Establish Clear Institutional Policies and Guidelines: Schools and districts should develop explicit policies regarding appropriate uses of generative AI for both teachers and students, addressing academic integrity expectations, privacy protections, acceptable use cases, and disclosure requirements.
3. Redesign Assessment Practices: Educational institutions should systematically revise assessment practices to emphasize higher-order thinking, creativity, and authentic application that AI cannot easily replicate, including performance-based assessments, portfolios, and projects.
4. Create AI Literacy Curricula: Schools should integrate AI literacy across grade levels, teaching students to evaluate AI outputs critically, understand AI limitations, and use AI ethically and responsibly, including explicit instruction in academic integrity.
5. Ensure Equitable Access and Establish Teacher Learning Communities: Educational institutions must address digital divide issues by ensuring equitable access to technology and AI tools, while also facilitating professional learning communities where teachers can share experiences and successful integration strategies.

6.2. Recommendations for Future Research

1. Longitudinal Studies of AI Integration: Future research should track teachers' perceptions, practices, and student outcomes as AI integration evolves over time, providing insights into how perceptions change with experience and what factors predict successful long-term integration.
2. Comparative and Cross-Cultural Studies: Research should compare teachers' perceptions and AI integration practices across different cultural, educational, and socioeconomic contexts to understand how contextual factors influence AI adoption and effectiveness.
3. Student Perspectives and Learning Outcomes: Future research should investigate students' perspectives on AI use in education and examine relationships between AI integration and student learning outcomes, engagement, and skill development across different educational levels.
4. Assessment Validity and Equity in AI-Enhanced Environments: Research should investigate how traditional assessment practices perform in AI-enhanced environments and examine how AI integration affects educational equity, developing strategies to ensure equitable access and benefit across diverse student populations.
5. Teacher Role Evolution and Ethical Frameworks: Future studies should investigate how teachers' roles evolve as AI becomes more prevalent, examining new competencies needed and developing comprehensive ethical frameworks addressing privacy, bias, transparency, and accountability in educational AI use.

REFERENCES

Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. Journal of AI, 7(1), 52-62. https://doi.org/10.61969/jai.1337500

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Cabellos, B., Ruiz-Robledillo, N., Ferrer-Cascales, R., & Albaladejo-Blázquez, N. (2024). University teachers' beliefs about the use of generative artificial intelligence for teaching and learning. Frontiers in Psychology, 15, Article 1468900. https://doi.org/10.3389/fpsyg.2024.1468900

Celik, I. (2023). Towards intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. Computers in Human Behavior, 138, Article 107468. https://doi.org/10.1016/j.chb.2022.107468

Chung, M., Ko, E., & Kang, H. (2024). Exploring the perceptions of Chinese pre-service teachers on the integration of generative AI in English language teaching: Benefits, challenges, and educational implications. Education and Information Technologies. Advance online publication.

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. Innovations in Education and Teaching International, 61(2), 228-239. https://doi.org/10.1080/14703297.2023.2190148

Creswell, J. W., & Poth, C. N. (2018). Qualitative inquiry and research design: Choosing among five approaches (4th ed.). SAGE Publications.

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. International Journal of Educational Technology in Higher Education, 20(1), Article 22. https://doi.org/10.1186/s41239-023-00392-8

Echave, M. B., Echave, M. A., & Echave, M. C. (2024). Embracing generative AI in education: Exploring teachers' perceptions, practices, and potential impact. Jurnal Pendidikan Indonesia Gemilang, 4(2), 123-135. https://doi.org/10.53889/jpig.v4i2.458

Ellis, R. A., Bliuc, A. M., & Goodyear, P. (2025). Qualitatively different teacher experiences of teaching with generative artificial intelligence. International Journal of Educational Technology in Higher Education, 22(1), Article 5. https://doi.org/10.1186/s41239-025-00532-2

Escario, I., Rodríguez-Sánchez, J. L., & Valero-Gil, J. (2024). Impulsando la Educación Superior con IAGen: oportunidades y retos para docentes. REDU: Revista de Docencia Universitaria, 22(2), 45-62. https://doi.org/10.4995/redu.2024.22065

Javahery, H., Saeli, M., & Khoshsima, H. (2025). Exploring Iranian novice EFL trainees' perceptions of ChatGPT use for lesson planning through a critical digital literacy lens. Information Retrieval Journal, 28(1), 1-28. https://doi.org/10.1007/s10791-025-09720-0

Kaplan-Rakowski, R., Grotewold, K., Hartwick, P., & Papin, K. (2024). Generative AI and teachers' perspectives on its implementation in education. Journal of Interactive Learning Research, 35(3), 443-470.

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kim, J. (2025). Perceptions and preparedness of K-12 educators in adopting generative AI. Research in Learning Technology, 33, Article 3448. https://doi.org/10.25304/rlt.v33.3448

Kim, N. J., Belland, B. R., & Axelrod, D. (2022). Teacher's perceptions of using an artificial intelligence-based educational tool for scientific writing. Frontiers in Education, 7, Article 755914. https://doi.org/10.3389/feduc.2022.755914

Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. International Journal of Management Education, 21(2), Article 100790. https://doi.org/10.1016/j.ijme.2023.100790

Lincoln, Y. S., & Guba, E. G. (1985). Naturalistic inquiry. SAGE Publications.

Majeed, A., Rana, S., & Khurshid, K. (2024). Role of generative artificial intelligence (Gen AI) in academia: University teachers' perspective within Pakistani higher education context. Research Square. https://doi.org/10.21203/rs.3.rs-5016419/v1

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. Biochemia Medica, 22(3), 276-282. https://doi.org/10.11613/BM.2012.031

Merriam, S. B., & Tisdell, E. J. (2016). Qualitative research: A guide to design and implementation (4th ed.). Jossey-Bass.

Ministry of National Education. (2023). Digital transformation in education: 2023-2027 strategic plan. Republic of Turkey Ministry of National Education.

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. Teachers College Record, 108(6), 1017-1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x

Mollick, E. R., & Mollick, L. (2023). Using AI to implement effective teaching strategies in classrooms: Five strategies, including prompts. The Wharton School Research Paper. https://doi.org/10.2139/ssrn.4391243

Moorhouse, B. L., Yeo, M. A., & Wan, Y. (2024). The effects of generative AI on initial language teacher education: The perceptions of teacher educators. System, 123, Article 103290. https://doi.org/10.1016/j.system.2024.103290

Munawar, S., Noor, S., & Ehsan, N. (2024). ChatGPT in classrooms: Transforming challenges into opportunities in education. arXiv preprint. https://doi.org/10.48550/arxiv.2405.10645

Nyaaba, M. (2024). Transforming teacher education in developing countries: The role of generative AI in bridging theory and practice. arXiv preprint. https://doi.org/10.48550/arxiv.2411.10718

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Administration and Policy in Mental Health and Mental Health Services Research, 42(5), 533-544. https://doi.org/10.1007/s10488-013-0528-y

Patton, M. Q. (2015). Qualitative research & evaluation methods: Integrating theory and practice (4th ed.). SAGE Publications.

Pettersson, F., Olofsson, A. D., & Lindberg, J. O. (2024). Generative AI and teachers - For us or against us? A case study. Linköping Electronic Conference Proceedings, 208, 35-44. https://doi.org/10.3384/ecp208005

Rasool, S. F., Samma, M., Wang, M., Anjum, A., & Iqbal, J. (2025). Perceptions of generative AI in teaching and learning: UTAUT and TPACK model-based study of teachers and students. Acta Psychologica, 253, Article 105827. https://doi.org/10.1016/j.actpsy.2025.105827

Reis-Andersson, A. (2024). AI generative in Brazil's public schools: The teachers' perspective. Proceedings of the International Conference on AI Research, 4(1), 234-248. https://doi.org/10.34190/icair.4.1.3219

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), 342-363. https://doi.org/10.37074/jalt.2023.6.1.9

Ruediger, D., Scherer, J., & Goben, A. (2024). Generative AI and postsecondary instructional practices: Findings from a national survey of instructors. EDUCAUSE Research Report. https://doi.org/10.18665/sr.320892

Sagocsoc, J. M., Sagocsoc, J. R., & Sagocsoc, M. T. (2025). Pros and cons of generative artificial intelligence in teaching-learning process: A sequential explanatory design. Premier Education and Management Journal, 1(1), 1-18. https://doi.org/10.70838/pemj.450705

Saud, M. S. (2025). Generative AI in English language teaching: Opportunities and challenges. Journal of Journalism and Mass Communication, 2(1), 45-58. https://doi.org/10.3126/jjmc2.v1i1.81437

Selwyn, N. (2019). Should robots replace teachers? AI and the future of education. Polity Press.

Soleimani, E., Ismail, L., & Mattar, J. (2025). Educators' perceptions of generative AI: Investigating attitudes, barriers and learning needs in higher education. Innovations in Education and Teaching International, 62(3), 456-471. https://doi.org/10.1080/14703297.2025.2530767

Tongchai, A., Suwannatthachote, P., & Nilsook, P. (2024). Thai teachers' perceptions of integrating generative AI in K-12 education: Opportunities and challenges. International Journal of Information and Education Technology, 14(8), 1089-1098.

Ututalum, C. (2025). Transformative potential of generative AI in higher education teaching practices. Environment and Social Psychology, 10(9), 2156-2170. https://doi.org/10.59429/esp.v10i9.3291

Van Manen, M. (2016). Researching lived experience: Human science for an action sensitive pedagogy (2nd ed.). Routledge.

Wang, L. (2024). Exploring teachers' perceptions of generative AI (ChatGPT) use on undergraduate teaching in EMI context: A literature review. Lecture Notes in Education Psychology and Public Media, 17, 234-245. https://doi.org/10.54254/2753-7048/2024.bo17885

Wardat, Y., Tashtoush, M. A., AlAli, R., & Jarrah, A. M. (2025). Teachers' use of generative AI in Jordanian universities: Practices and perceptions. Forum for Linguistic Studies, 7(6), 2345-2361. https://doi.org/10.30564/fls.v7i6.9644

Yılmaz, A., & Kaya, M. (2024). Digital transformation in Turkish education: Challenges and opportunities. Turkish Journal of Educational Technology, 24(3), 112-128.

Yue, M., Jong, M. S. Y., & Dai, Y. (2025). Obstacles or opportunities: Teachers' concerns about adopting generative AI in learning and teaching. Proceedings of the 2025 International Conference on Advanced Learning Technologies, 512-516. https://doi.org/10.1109/icalt64023.2025.00084

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), Article 39. https://doi.org/10.1186/s41239-019-0171-0

The Impact of Social Media Users' Gratification Seeking on Usage Frequency: The Case of Instagram a Quantitative Research

Bilge YÜCE ABAY
PhD Student, Department of Communication Sciences, Institute Of Social Sciences, Sakarya University. Turkiye
bilge.abay@ogr.sakarya.edu.tr
ORCID: 0009-0008-7586-2533

Prof. Dr. Aytekin İŞMAN
SAU, Faculty of Communication
isman@sakarya.edu.tr

ABSTRACT

With technological advancements, social media platforms have become powerful communication tools that satisfy users' needs for information, entertainment, and socialization (Kaplan & Haenlein, 2010). Among these platforms, Instagram stands out in brands' digital communication strategies due to its structure and the fact that its users produce content on different themes (Saatçioğlu & İnanç, 2020). The main objective of this research is to systematically and comprehensively reveal the extent to which Instagram users' various gratification-seeking behaviors affect their frequency of platform use.
The study utilized a descriptive survey model, a quantitative research method. Data were collected through an online survey technique, and 388 students from Bilecik Şeyh Edebali University were included. The Instagram Usage Motivations Scale, adapted into Turkish by Saatçioğlu and İnanç (2020), was used as the measurement tool in the study. The scale consists of five sub-dimensions and a total of 28 items. In addition to the scale, questions regarding participants' demographic characteristics and frequency of platform use were included in the data collection form. In the reliability analysis of the scale, the Cronbach's Alpha value for the entire scale was determined as α=.824; the alpha coefficients of the sub-dimensions were reported as α=.828 for social interaction, α=.766 for archiving, α=.792 for self-expression, α=.732 for escape, and α=.725 for surveillance, meeting the threshold of .70 for internal consistency, which is required for reliability. SPSS 31.0.1 software was used to analyze the data, and descriptive statistics, T-test, and ANOVA analyses were applied.
The findings yield empirical evidence showing that the frequency of Instagram use is a determining factor, particularly in terms of social interaction, archiving, and surveillance gratification. The findings have demonstrated that types of gratification differ as usage intensity increases, providing theoretical and practical contributions to the social media literature and digital communication strategies
Keywords: social media, instagram, seeking gratification, uses and gratifications theory, frequency of use

INTRODUCTION

The rapid development of digital communication technologies and the penetration of the internet into every aspect of social life have led to profound changes in a wide range of areas, from how individuals communicate and access information to their understanding of entertainment and consumption behaviors. Unlike the one-way and limited interaction structure of traditional media, social media platforms enable active user participation, content creation, and real-time interaction (Bat & Vural, 2010). In this context, the focus is on what individuals do with the media rather than what the media does to individuals; in the digital age, users actively prefer social media platforms for motivations such as acquiring information, social interaction, and entertainment (Cevher & Ustakara, 2019).

Interactions on platforms become part of users' organic content experiences and offer a higher potential for interaction (Saatçioğlu & İnanç, 2020). When Instagram, the subject of this research, is examined within the framework of Uses and Gratifications Theory, the frequency of users' use of the platform and their search for gratifications such as social interaction, self-expression, surveillance, and escapism constitute an important area of research in understanding the dynamics of social media use.

SOCIAL MEDIA

Social media is a broad digital interaction environment encompassing various communication tools and platforms based on advancements in internet technologies (Okmeydan, 2020). It consists of interactive platforms that utilize mobile and web-based technologies to enable individuals and communities to share, co-create, discuss, and modify user-generated content (Kietzmann et al., 2011). The multifaceted nature of these platforms allows users to develop different usage patterns according to their cognitive, emotional, and social needs. The variety of content

offered directly affects the time users spend on the platform, their level of interaction, and their motivation for use.

On some platforms, individuals move from being passive viewers to becoming content producers and active participants; thus, social media takes on a more central role in daily life practices (Cevher & Ustakara, 2019). These usage patterns form an important basis for understanding the types of gratification users aim to obtain from social media. Indeed, social media users' platform preferences, interaction styles, and content consumption habits are shaped by their individual motivations and pursuit of gratification (Özeltürkay & Yarımoğlu, 2021). In this context, high participation and continuous content production are among the key dynamics determining the frequency and duration of social media platform use. This process provides a foundation for individuals to express themselves and meet their various social needs (Kietzmann et al., 2011).

INSTAGRAM

Instagram, creating a space where visual communication dominates the social media environment, stands out from other networks with its unique structure (Saatçioğlu & İnanç, 2020). Its photo and video-based structure allows users to share their experiences aesthetically and gain social approval; therefore, the platform creates a strong motivation for use, especially among young users (Aktan, 2018). In addition, Instagram increases user engagement through its ability to share content simultaneously across various platforms and its hashtag/tagging features (Türkmenoğlu, 2014). The platform increases active usage time by stimulating users' desire for social interaction and visibility (Diker & Gencer, 2019). At the same time, Instagram's dynamic structure supports the archiving instinct, which is the desire of users to preserve their experiences (Uzun, 2019). Users not only gain likes but also share their products and experiences with their followers by creating content in blog and vlog formats (Kurtulan, 2017).

On the other hand, social media is generally one of the areas where surveillance is most intense (Aka, 2024). Instagram's constantly evolving content feed allows users to spend more time on the platform, distancing themselves from daily life and indirectly supporting a motivation for escapism. Users constantly satisfy their need for information by monitoring events around them through the platform, thus supporting the surveillance function (Güleç & Köker, 2021). In this context, Instagram, with its visual and interaction-oriented structure, provides a dynamic environment that intensifies platform usage by activating users' motivations for social interaction, visibility, and content creation (Kietzmann et al., 2011). Therefore, the types of gratification offered by Instagram constitute a critical context for examining the effects of users' usage frequency within the framework of Uses and Gratifications Theory (Cevher & Ustakara, 2019).

THE SEARCH FOR FULFILLMENT

Social media platforms have become an indispensable part of the daily routines of billions of users today. A study by Alhabash et al. (2024) revealed that Instagram showed a significant advantage over other social media platforms in terms of gratification seeking among university students. In this context, social media has become a space where users not only communicate but also constantly seek gratification. This quest leads users to visit and interact with the platform regularly (Hoque & Hossain, 2023). Social media platforms prevent users from completing their quest for satisfaction as a one-time process by offering instant notifications, a constantly updated stream of content, and opportunities for interaction (Cevher & Ustakara, 2019). Instead, the platform creates a cycle of satisfaction by constantly encouraging users to interact with new content (Saatçioğlu & İnanç, 2020).

Users may create content driven by the desire to build social connections and gain validation, while they may also spend time on the platform for reasons such as entertainment, acquiring information, or escaping daily stresses. These types of satisfactions provide an important framework for understanding which content users gravitate towards, which posts they react to, and their interaction patterns on the platform (Uzun, 2019). At this point, Uses and Gratifications Theory offers a strong theoretical framework for systematically explaining users' motivations for using social media and the gratifications they obtain (Yıldırım et al., 2018). This theory provides a basis for understanding which types of gratification users are seeking when they turn to the platform more frequently, and how these gratification pursuits affect the frequency of use (Uzun, 2019).

Research conducted specifically on Instagram within the framework of Uses and Gratifications Theory shows that users utilize the platform in pursuit of various gratifications. In their pioneering study examining Instagram usage motivations, Lee et al. (2015) identified five key dimensions of gratification. Basic motivations are classified as social interaction, archiving, self-expression, escapism, and surveillance. This study has become one of the fundamental references for Instagram research. Saatçioğlu and İnanç (2020), in their research conducted in the context of Turkey, examined Instagram usage motivations in similar categories and investigated the relationship

between these motivations and user satisfaction**.** These five dimensions of satisfaction are used in the literature as a fundamental framework for understanding Instagram user behavior.

Social interaction refers to users' desire to connect with others and maintain social relationships; archiving refers to the tendency to preserve memories and create digital diaries (Sheldon & Bryant, 2016). Self-expression reflects users' desire to present their identities and seek social approval; while surveillance reflects the urge to monitor and gather information about the lives of others (Bilgici Oğuz & Atasoy, 2018). The escapist motive has been found to be associated with the tendency to use Instagram as a means of distancing oneself from real life (Kırcaburun & Griffiths, 2019). In this context, understanding which gratification-seeking behaviors of Instagram users increase their frequency of platform use is important for discussing the validity of Uses and Gratifications Theory in contemporary digital communication environments. Therefore, this study aims to examine the relationship between Instagram users' gratification seeking and their frequency of platform use within the framework of Uses and Gratifications Theory.

PURPOSE

The main objective of this research is to examine the differences in the gratification-seeking behaviors of Instagram users based on their frequency of platform use, within the framework of Uses and Gratifications Theory. New media technologies have eliminated the one-way nature of communication tools, accelerated the increase in information, and fundamentally changed the way individuals use media, affecting their daily lives (Alioğlu, 2015; Çömlekçi & Başol, 2019). On the other hand, research shows that the younger generation uses digital-based media intensively and conducts most of their daily life practices through these tools (Özdemir, 2021). Social media offers individuals the opportunity to introduce themselves and interact with others; it enables them to make their daily experiences visible through functions such as location sharing, commenting, and creating stories (Salderay & Erten, 2021). The different dimensions of satisfaction derived from Instagram, such as social interaction, self-expression, surveillance, archiving, and escapism, appear to determine users' orientations and usage practices on the platform. The instant feedback mechanisms offered by the platform reinforce these orientations, contributing to the formation of sustainable usage habits. This research has three sub-objectives.

The primary objective of this research is to determine the level of difference in the frequency of Instagram users' use of five basic types of gratification (social interaction, self-expression, surveillance, archiving, and escapism). According to Uses and Gratifications Theory, individuals consciously choose and use social media platforms to meet their own needs. Especially on visually oriented platforms like Instagram, the feedback users receive from their posts (likes, comments, follower increase) reinforces their search for gratification, leading to more intense engagement on the platform (Sheldon & Bryant, 2016). In this context, identifying the differences in the frequency of use of these types of gratification is important for understanding the motivations behind users' engagement with the platform and how these motivations shape their usage behavior.

The second objective of the study is to determine whether the pursuit of gratification differs significantly among the five basic types of gratification based on frequency of use. This objective is structured based on Uses and Gratifications Theory. The literature indicates that types of gratification, such as social interaction and self-expression encourage users to share and interact more on the platform (Sheldon & Bryant, 2016). In contrast, it is stated that types of gratification such as surveillance and escapism lead users to spend more time on the platform, but in a more passive manner (Aktan, 2018). However, the limited number of studies on which of these gratification dimensions most strongly predicts the frequency of Instagram use makes this research important. Accordingly, comparing which type of gratification users prioritize more and the relative effects of these gratifications on the frequency of use will significantly contribute to understanding the dynamics of social media use.

The third objective of this study is to reveal the differences in gratification seeking and usage frequency among Instagram users based on demographic characteristics such as gender, daily usage time, and education level. The literature frequently emphasizes that social media usage behaviors differ significantly according to demographic variables. For example, it is stated that gender plays a decisive role in social media usage motivations, with female users tending more towards social interaction and self-expression gratification, while male users are motivated by surveillance and information gathering (Sheldon & Bryant, 2016). In terms of daily platform usage time, it is stated that heavy users turn to the platform more frequently in search of escapism and entertainment, while light users enter the platform more for the purposes of establishing social connections and acquiring information (Saatçioğlu & İnanç, 2020). The education level variable shapes users' content consumption patterns and platform usage purposes; it is observed that users with higher education levels use Instagram for purposes such as acquiring information and professional networking, while users with lower education levels prioritize entertainment and

escapism (Özdemir, 2021). It is noted that a significant portion of existing studies on social media usage in Turkey address these differences mostly through descriptive analyses (Alioğlu, 2015; Çömlekçi & Başol, 2019).

In this study, these differences were examined through a systematic analysis of the differences in the frequency of use of Instagram in relation to users' pursuit of gratification. The research was conducted with university students who use Instagram, and the Instagram Usage Motivations Scale, adapted into Turkish by Saatçioğlu and İnanç (2020), was administered to the participants. Based on the findings, the differentiation levels of five basic gratification types according to their frequency of use were revealed, and the changes in gratification seeking according to demographic variables were determined. The study results showed that Instagram users differed significantly in their frequency of use, particularly with regard to gratification types such as surveillance, archiving, and social interaction; these findings offer important academic implications regarding the validity of social media use in current digital communication environments within the framework of the theory.

IMPORTANCE

The fact that social media platforms have become an integral part of daily life makes understanding the gratifications users seek in these environments crucial, both individually and socially. Social media has become a fundamental tool for individuals to meet various needs such as social interaction, information gathering, entertainment, and self-expression (Bhatiasevi, 2024). However, the unconscious and uncontrolled use of social media can lead to negative consequences such as addiction, decreased academic performance, and psychological problems (Masrom et al., 2023). This research reveals the effects of Instagram users' gratification-seeking on their frequency of use, offering important findings for individuals to regulate their social media experiences in a healthier and more conscious way. Furthermore, the study aims to contribute to social media research from both theoretical and practical perspectives by systematically examining Instagram usage within the framework of Uses and Gratifications Theory. This research has three important aspects.

Firstly, the research reveals the underlying motivations of social media use behavior by examining the differences in the frequency of Instagram users' use of the platform in relation to their pursuit of gratification. It is known that usage motives and the gratifications sought increase the intensity of social media use (Hoque & Hossain, 2023). However, the allure and easy accessibility of social media can lead to various negative consequences for individuals, especially young users (Koç, 2023). This situation necessitates that users become aware of the purposes for which they use platforms and structure their social media experiences in a more conscious and purposeful way (Yolcu & Çiftçi, 2023). This research aims to identify which needs are met through Instagram, acknowledging that individuals are not passive consumers but active users who make choices based on their needs on social media (Biliciler, 2018). Accordingly, the study aims to develop a framework that will help individuals understand their motivations and usage intensity for social media use.

Secondly, this research offers a methodological contribution to understanding social media usage dynamics by identifying which of the five basic types of gratification has a stronger influence on the frequency of Instagram use. Individuals' motivations for using social media affect the frequency and intensity of their visits to the platform; moreover, these motivations change as the frequency of use increases (Whiting & Williams, 2013). For example, problematic social media users tend to show higher propensity for certain types of motivation, such as escapism, entertainment, social interaction, and instant gratification (Masrom et al., 2023). This research compares the relative effects of five types of gratification, as well as examining differences in gratification pursuits based on demographic variables such as gender, daily usage time, and education level. The findings help users understand which motivations they prioritize, enabling them to manage their social media experiences in a more conscious, healthy, and balanced way.

Thirdly, the research provides a theoretical contribution to social media research by testing Uses and Gratifications Theory in the context of Instagram. Uses and Gratifications Theory is one of the theoretical frameworks widely used in social media research because it views individuals not as passive consumers of media, but as users who make active choices to satisfy their needs (Bhatiasevi, 2024). However, the fact that each social media platform has different characteristics and dynamics indicates that user motivations and gratification pursuits can vary from platform to platform (Sheldon & Bryant, 2016). Instagram, with its focus on visual content sharing, distinctly differs from text-based platforms like Facebook and Twitter, and this difference influences users' pursuit of gratification. The literature indicates that Instagram's visual and aesthetic nature particularly highlights archiving and surveillance gratification (Ürkmez & Eskicumalı, 2021). While interest in Instagram's social, cultural, and psychological impacts is growing, comprehensive analyses of the gratification users derive from the platform and how these gratifications shape usage frequency are limited. The scarcity of studies comparing the five basic types of gratification, in particular, makes it difficult to comprehensively understand user behavior. This study examines the effects of five gratifications on Instagram usage frequency in a Turkish sample, aiming to both test the validity

of the theory on visually oriented platforms and fill a gap in the literature. Furthermore, the findings are not limited to Instagram alone, but provide a foundation for understanding usage dynamics based on visual and content sharing across various platforms. In this context, the research offers a theoretical and empirical basis for future studies in the fields of digital media psychology and social media literacy.

Based on these three fundamental points, this research aims to contribute to a better understanding of social media usage motivations at both academic and societal levels. The findings will help users manage their Instagram experiences more consciously, while also providing a foundation for new studies on digital media literacy. Thus, the study serves as an important resource that addresses social media usage satisfaction within a holistic framework.

THEORETICAL FRAMEWORK

The theoretical basis of this research is Uses and Gratifications Theory. Developed in the early 1940s by researchers who focused on audience gratification and argued that the influence of media is limited, Herta Herzog, Bernard Berelson, Joseph Klapper, Elihu Katz, and Paul Lazarsfeld laid the foundation for this approach with the idea that "Media does not directly do anything to people; what matters is what people do with the media" (Erdoğan & Alemdar, 2010). Uses and Gratifications Theory was developed as an alternative to traditional media influence approaches, viewing the user not as a passive recipient but as an active, conscious, and purposeful individual in media use (DeFleur, 2016). According to this theoretical framework, viewers freely choose the media content that best meets their needs. In this regard, the fundamental characteristics of Uses and Gratifications Theory, which treats the viewer as an active, needs-oriented, conscious, purposeful, and responsible participant, offer a critical framework for explaining social media use and users' pursuit of gratification (Küçükkurt et al., 2016). The five core characteristics of Uses and Gratifications Theory are as follows:

The Audience is Active: The most fundamental assumption of Uses and Gratifications Theory is that the audience is active in their relationship with the media. The individual is not in a passive position but becomes active by initiating their own interaction with the media. Unlike classical approaches that accept the passivity of the receiver, the theory supports the view that users are active in using the media according to their own needs (Küçük Durur & Akbaba, 2022).

Media Use is Purposeful and Driven by Needs: Media use is not random, but serves a specific purpose. Individuals' media choices stem from motivations aimed at fulfilling basic socio-psychological needs. The fundamental premise of the Uses and Gratifications Approach is to determine how individuals consciously use mass media to satisfy their needs (Oral, 2022).

Choice Based on Need Satisfaction: Viewers are aware of which media type or content best meets their needs and make their choices accordingly; thus, they freely and consciously select the media and programs that will best satisfy their requirements (Küçükkurt et al., 2016).

Media is a source chosen from among alternatives: Media is not the only way to satisfy needs. Viewers can achieve the same satisfaction (e.g., relaxation, socialization) by consciously choosing and directing their media use according to their own needs. This assumption is strengthened in the current context by the fact that media platforms offer viewers multiple content and alternative choices, providing opportunities for diverse preferences and individual satisfaction (Küçük Durur & Akbaba, 2022).

The Value of Media Content is Not Judged: The theory does not judge the cultural quality, moral value, or importance of media content. The focus is not on the content itself, but on the gratifications obtained, the satisfactions that viewers actually experience by using a particular medium. In this context, instead of focusing on the content, it concentrates particularly on the needs, motives, expectations, and preferences of the audience (Quan-Haase & Young, 2010). The need to meet people's needs and to discover the tools to support these needs constitutes the fundamental motivation of media use research (Akar, 2023; Quan-Haase & Young, 2010). In this context, it has been hypothesized that individuals who resist influence will consciously choose to use communication tools that they believe best meet their needs and expectations (Akar, 2023). Accordingly, individuals' conscious choices in media use guide their processes of satisfying their needs (DeFleur, 2016). It is assumed that the types of gratification that influence the frequency of use of a particular mass communication tool are a fundamental element in motivating individuals and determining the frequency of use (Oral, 2022). In particular, the purposeful and needs-oriented nature of media use is important in explaining the frequency of use and content preferences of Instagram users, who are the subject of our research (Oral, 2022; Küçükkurt et al., 2016).

This study was conducted within the framework of the theory's five fundamental characteristics. The first fundamental characteristic is the active participation of the audience; users consciously select and use media according to their own needs and purposes. Within the scope of the study, it was determined and interpreted that Instagram users utilize the platform in this manner. The second fundamental characteristic is that media use is purposeful and driven by needs; The study examined users' diverse pursuit of satisfaction (dependent variables) and frequency of use (independent variable). A third key characteristic is that users make their media choices based on their satisfaction expectations to meet their own needs. The study analyzed the frequency of Instagram users' use according to their content choices based on their types of satisfaction. A fourth key characteristic is that media is a source chosen from among alternatives; this allowed for the evaluation of users' pursuit of satisfaction within the framework of their usage frequency. A fifth characteristic is that the value of media content is not judged; in this study, users' satisfaction with Instagram use was evaluated independently of the type and quality of the content. Theory was used as the fundamental guide shaping the conceptual framework and data analysis of the research.

LITERATURE REVIEW

This literature review examines the varying degrees of difference in social media users' gratification-seeking behaviors based on their frequency of use and reveals trends, findings, and gaps in the existing literature. Ten articles published in scientific journals on the subject were reviewed.

Firstly, the study conducted by Whiting and Williams (2013) aims to reveal the functionality of Uses and Gratifications Theory in the field of social media and to examine the gratifications users derive from these platforms. Through in-depth interviews with 25 users, ten different gratification factors motivating social media use were identified in order to provide a comprehensive framework for understanding why individuals use social media. These factors were classified as social interaction, information seeking, spending time, entertainment, relaxation, communicative benefit, convenience benefit, expressing opinions, sharing information, and observing others. In particular, social interaction, mentioned by 88% of participants, and information seeking, highlighted by 80%, emerged as the most dominant motivations for use and directly triggered the frequency of use. The findings of this research provide a strong theoretical foundation for our study, which examines the impact of social media users' gratification-seeking on their usage frequency using Instagram as an example. Understanding how these ten different types of gratification, as defined in the literature, shape user motivations helps us to gain a deeper understanding of the mechanisms that trigger usage intensity in our research.

Secondly, the study conducted by Başoğlu and Yanar (2016) aims to examine the purposes of social media use and habits of university students within the framework of Uses and Gratifications Theory. The survey, conducted on 423 students, revealed that participants primarily used social networks for entertainment, information gathering, and communication. Validity and reliability analyses of the scale used in the research demonstrated that young adults use social media not only to pass the time but also to maintain interaction, a social need. It has been highlighted that entertainment and social interaction satisfactions are key factors determining users' daily usage routines and the amount of time they spend on the platform. The findings of this research provide an application-level basis for our study, which examines the impact of social media users' satisfaction-seeking on their usage frequency using Instagram as an example. The usage motivations of university students help us to understand more deeply the mechanisms that trigger usage intensity in our research.

Thirdly, the study conducted by Ürkmez and Eskicumalı (2021) aims to examine individuals' motivations for using Instagram during the COVID-19 pandemic within the context of Uses and Gratifications Theory. The research, conducted under the isolation and curfews imposed during the pandemic, revealed that needs such as accessing information, entertainment, and social interaction directly shaped Instagram usage during this period. The study highlights how limitations in the physical world have transformed into a more intense pursuit of gratification in the digital world. It specifically notes that individuals isolated from the outside world are turning to Instagram more than ever to satisfy their need for social interaction and leisure time. The findings of this research provide a current and platform-focused perspective for our study, which examines the impact of social media users' gratification-seeking behaviors on usage frequency through the example of Instagram. The critical role of Instagram-specific gratification-seeking behaviors during times of crisis helps us to understand more deeply the mechanisms that trigger usage intensity in our research.

Fourthly, the study conducted by Aslan (2018) aims to examine the Instagram use of university students and the levels of satisfaction they derive from this use, based on Uses and Gratifications Theory. The empirical study, conducted on students of the Faculty of Communication at Erciyes University, revealed that demographic variables such as gender, education level, and department significantly affected the levels of satisfaction obtained from Instagram. The research demonstrates how users' motivations for using the platform can be categorized

according to personal characteristics. It particularly highlights that female students and graduate students derive significantly higher satisfaction from using Instagram compared to other groups. The findings of this research provide a basis for demographic comparison in our study, which examines the impact of social media users' pursuit of satisfaction on their frequency of use, using Instagram as an example. Understanding how demographic variables differentiate satisfaction processes helps us to gain a deeper understanding of the mechanisms that trigger usage intensity in our research.

Fifthly the study conducted by Ting et al. (2015) aims to examine users' core beliefs about Instagram use and their attitudes towards the platform in light of Uses and Gratifications Theory. Using an exploratory research design, the study found that social networks have become an indispensable method of communication in the modern world and that usage is fueled by positive beliefs towards the platform. The study discusses how technological advancements influence users' media choices. In particular, the trust and aesthetic perception associated with the platform's visual structure have been highlighted as reinforcing users' intentions to consistently use Instagram. The findings of this research provide a perceptual framework for our study, which examines the impact of social media users' gratification-seeking on their usage frequency using Instagram as an example. Users' cognitive beliefs about the platform help us to understand more deeply the mechanisms that trigger usage intensity in our research.

Sixthly, the study conducted by Çulfacı, and Kılıçhan (2023) aims to examine the relationship between impression management behavior, Instagram usage intensity, and food photography sharing (foodstagramming) within the framework of Uses and Gratifications Theory. Based on data obtained from 407 participants, a linear relationship was found between Instagram usage intensity and the motivation to leave a positive impression on others. The research analyzes how sharing visual content is combined with the need for social approval. It has been suggested that the gratification of receiving social approval and self-presentation directly increases the frequency of Instagram use and the intensity of content sharing. The findings of this research provide behavioral support for our study, which examines the impact of social media users' gratification-seeking on their usage frequency using Instagram as an example. The process by which psychological gratification translates into the tangible action of usage frequency helps us to understand more deeply the mechanisms that trigger usage intensity in our research.

Seventhly, the study conducted by Simamora, and Andika (2019) aims to examine the Instagram usage intensity and usage patterns of Generation Z middle school students within the context of Media Literacy and R&D. Using a qualitative approach, the study found that young users have made Instagram an integral part of their daily lives and exhibit a very high usage frequency. The study describes the digital habits of this group, referred to as the "now generation." The desire for instant sharing and constant interaction with their social circle has emerged as the most significant factor driving usage intensity, particularly among Generation Z. The findings of this research provide a generational data source for our study, which examines the impact of social media users' pursuit of gratification on their usage frequency using Instagram as an example. The platform usage patterns of young users help us to gain a deeper understanding of the mechanisms driving usage intensity in our research.

Eighthly, the study conducted by Özdemir Süzer (2024) aims to examine the reasons for Instagram users' satisfaction and their behavioral intentions after satisfaction, focusing on Uses and Gratifications Theory. The research, conducted on a Generation Z sample, revealed that the dimensions of entertainment, utility, and socialization have a decisive effect on user satisfaction and brand attitude. The study emphasizes that the satisfaction obtained is not limited to platform use but extends to the intention to purchase. Specifically, the feeling of satisfaction derived from the psychological fulfillment provided by the platform has been shown to increase users' time spent on Instagram and the frequency of their interactions. The findings of this research provide an outcome-oriented model for our study, which examines the impact of social media users' pursuit of satisfaction on their usage frequency using Instagram as an example. The transformation of this satisfaction into a sustained usage behavior helps us to understand more deeply the mechanisms that trigger usage intensity in our research.

Ninthly, the study conducted by Taş, Özen and Bölen (2021) aims to examine the cognitive, emotional, and behavioral factors influencing attitudes towards using Instagram in relation to Uses and Gratifications Theory. The study found that the attitudes users develop towards Instagram are fundamentally based on their expectations of the visual aesthetics and emotional gratification offered by the platform. The research analyzes how these attitudes solidify media usage habits. In particular, emotional attachment, visual appeal, and the perceived benefit derived from the platform have been highlighted as the most critical factors reinforcing usage frequency. The findings of this research provide a cognitive-emotional framework for our study, which examines the impact of social media users' gratification-seeking on their usage frequency using Instagram as an example. The process by which user attitudes translate into action and usage frequency helps us to understand more deeply the mechanisms that trigger usage intensity in our research.

Tenthly, the study conducted by Aksoy and Allahverdi (2025) aims to examine the systematic relationship between social media use motivations and user engagement within the framework of Uses and Gratifications Theory. As a result of a systematic review including numerous studies, 19 different types of motivation explaining users' reasons for choosing social media were identified. This study synthesizes current literature to reveal how user characteristics determine their pursuit of gratification. It has been highlighted that this diversity of motivations directly influences both platform preference and the intensity and frequency of interaction within applications. The findings of this research provide comprehensive and up-to-date literature support for our study, which examines the impact of social media users' pursuit of gratification on their frequency of use, using Instagram as an example. The diversity of new types of gratification helps us to understand more deeply the mechanisms that trigger usage intensity in our research.

METHODS

This study is designed using a quantitative research approach to examine the differences in the frequency of use of social media users in relation to their pursuit of gratification. Quantitative research is based on the assumption that social phenomena can be measured and explained with numerical data; this approach allows for the systematic analysis of relationships between variables using statistical methods (Robson & McCartan, 2016). To achieve the study's objective and answer the research questions, a descriptive survey model was used in accordance with the data collection techniques and analysis processes that constitute the methodological framework of the research. This model, which aims to systematically define and describe an existing situation, objectively reveals the current state of individuals' attitudes, behaviors, and characteristics (Karasar, 2012). Through this model, the effect between social media users' pursuit of gratification and their frequency of use is examined. In the model used to determine the sample's behavior, the researcher does not intervene in the survey but reveals the existing situation (Çalhan &Uslu, 2025).

In this context, the procedures adopted to ensure the validity and reliability of the scale used in the study, and the steps followed in the application phases, are presented systematically. To ensure the correct application of the chosen method, the research literature, previous studies, and the specific requirements of the study were examined. The theoretical basis for explaining the frequency of use in the context of gratification seeking in this research is the Uses and Gratifications Theory. The theory considers individuals' media use not as accidental, but as a purposeful process shaped by specific needs and motivations, and evaluates users' content choices as those of active subjects (Küçükkurt et al., 2016). Within this scope, in the research;

Independent Variables

Instagram usage frequency is measured through demographic questions that determine how long participants use the platform daily and how many years they have been active. Through this data, Instagram usage duration is analyzed according to participants' pursuit of satisfaction.

Dependent Variables

Instagram users' pursuit of fulfillment; users' motivations for using Instagram are measured within the scope of their pursuit of fulfillment, which includes individuals' diverse social needs and expectations.

Problem Sentence

Do Instagram users' diverse pursuits of satisfaction play a significant role in determining their frequency of platform use?

Sub-problems and hypotheses

This study was conducted in line with the determined sub-problems and hypotheses; it aims to examine in detail the differences in the levels of satisfaction with social interaction, archiving, self-expression, escapism, and surveillance on the frequency of Instagram use.

In the study, the sub-problems and hypotheses were determined as follows:
Sub-Problem 1. Is there a significant difference between social interaction satisfaction and the frequency of Instagram use?
H1: There is a significant difference between social interaction and the frequency of Instagram use.
H01: There is no significant difference between social interaction and the frequency of Instagram use.

Sub-Problem 2. Is there a significant difference between archiving satisfaction and the frequency of Instagram use?
H2: There is a significant difference between archiving and the frequency of Instagram use.
H02: There is no significant difference between archiving and the frequency of Instagram use.

Sub-Problem 3. Is there a significant difference between self-expression satisfaction and the frequency of Instagram use?
H3: There is a significant difference between self-expression and the frequency of Instagram use.
H03: There is no significant difference between self-expression and the frequency of Instagram use.

Sub-Problem 4. Is there a significant difference between escapism and the frequency of Instagram use?
H4: There is a significant difference between escapism and the frequency of Instagram use.
H04: There is no significant difference between escapism and the frequency of Instagram use.

Sub-Problem 5. Is there a significant difference between surveillance gratification and the frequency of Instagram use?
H5: There is a significant difference between surveillance and the frequency of Instagram use.
H05: There is no significant difference between surveillance and the frequency of Instagram use.

Research Model

This study employed a descriptive survey model to examine the current state and differences in the pursuit of satisfaction among Instagram users in terms of platform usage frequency and demographic variables. Descriptive survey models are approaches that aim to describe a past or present situation as it is (Büyüköztürk, 2022). Accordingly, the study tested whether there was a statistically significant difference in the pursuit of satisfaction based on the independent variables identified. In the analysis of the data, the characteristics of the participants were described; and parametric tests were used for intergroup comparisons. A t-test was used to compare the means of two independent groups, and one-way analysis of variance (ANOVA) was applied to examine the differences between the means of three or more groups (Mishra et al., 2019; Karadavut, 2021). In this respect, the research has a descriptive nature, aiming to reveal users' pursuit of satisfaction in terms of various variables.

Population

The population of this research consists of 20,600 students studying at Bilecik Şeyh Edebali University during the 2024-2025 academic year.

Sample

This study sample consists of 388 students from Bilecik Şeyh Edebali University who actively use Instagram. Probability-based random sampling was used to select the sample. The fundamental principle behind this method is that each unit in the population has an equal probability of being selected, thereby minimizing bias (Erdoğan, 2012).

Time, cost, and access conditions were considered in determining the sample size. Furthermore, the inclusion of participants from different departments and class levels increased sample heterogeneity and contributed to the reliability of the analysis results. The fact that the participants actively use Instagram strengthens the validity of the study in terms of evaluating their pursuit of satisfaction.

Data Collection Tool

This study utilized a quantitative research method, specifically the survey technique. Data were collected online via Google Forms. The Instagram Usage Motivations Scale, adapted by Saatçioğlu and İnanç (2020), was used in the research. The scale consists of 28 items across five sub-dimensions (Social Interaction, Archiving, Self-Expression, Escape, and Surveillance). The first section of the scale includes questions about participants' demographic characteristics such as gender, age, and education level, as well as their frequency of use on the platform. The items in the scale were rated using a 5-point Likert scale (1=Strongly Disagree, 2=Disagree, 3=Undecided, 4=Agree, 5=Strongly Agree). A Cronbach Alpha coefficient of 0.70 or higher indicates the scale's reliability (Nunnally & Bernstein, 1994). The Cronbach's Alpha value of the scale used in this context is $\alpha = .824$, and the alpha coefficients of the sub-dimensions are reported as $\alpha=.828$ for social interaction, $\alpha=.766$ for archiving, $\alpha=.792$ for self-expression, $\alpha=.732$ for escape, and $\alpha=.725$ for surveillance, providing a value of .70 and above, which is necessary for reliability (Saatçioğlu & İnanç, 2020). Ethical approval for this study was obtained from the Sakarya University Social and Human Sciences Ethics Committee with decision number 93/35 dated 19/11/2025.

Data Analysis

The data obtained from the research were analyzed using SPSS 31.0.1, a statistical package for social sciences. First, to ensure the suitability of the dataset for analysis, missing data checks, outlier checks, and normality tests were performed. Descriptive statistics for the participants included percentages (%), frequencies (f), means ($\bar{X}$), and standard deviations; thus, demographic variables and satisfaction-seeking behaviors were reported in detail.

For the descriptive statistics of the participants, frequency (percentage) distributions were analyzed, and independent samples t-tests and one-way ANOVA analyses were conducted to examine significant differences according to group variables. In all statistical analyses, the significance level was accepted as $p < .05$, and the findings were evaluated in accordance with this criterion. In this context, the differences in Instagram users' satisfaction-seeking behaviors based on their frequency of platform use were revealed through statistical analyses.

Assumptions

The following assumptions were considered in this research:

- It is assumed that participants answered the survey questions honestly and objectively. For the reliability of survey studies, the degree to which the respondents provided accurate and reliable information is a fundamental requirement for the validity of the research results (Zırhlıoğlu, 2018).
- It is assumed that the selected sample represents the research population. In research, it is assumed that it represents the population to which generalizations will be made (Büyüköztürk, 2022).
- It is assumed that the scale items validly represent the constructs they are intended to measure. In validity studies, it is of great importance that the scales are consistent with the theoretical constructs they are intended to measure (DeVellis, 2016).
- It is assumed that the assumptions required by the applied statistical analyses are fully and accurately met. Homogeneity of variance is considered an important assumption in research in terms of the proper conduct of analyses and the reliability of the findings obtained (Çelikten Demirel, et al., 2025).
- It is assumed that participation in the research is entirely voluntary and that ethical principles are fully complied with at all stages. In studies conducted on participants, the elements of informed consent, understanding, and voluntariness must be fulfilled in order for the consent to be ethically and legally valid (Ersoy, 2015).
- It is assumed that the internal consistency levels of the scales will meet the accepted psychometric reliability criteria. A Cronbach Alpha coefficient of 0.70 and above is generally considered sufficient in terms of the internal consistency reliability of the scales (DeVellis, 2016).

Premises

- This research assumes that certain statistical and methodological assumptions are met. These assumptions are critical elements that directly affect the accuracy of the types of analysis used, the validity of the results obtained, and their reliability.
- Below, the basic assumptions of the research are explained within a theoretical framework:
- It is assumed that the data are normally distributed. The assumption of normality is one of the basic statistical conditions required for the application of parametric analyses (Robson & McCartan, 2016).
- It is assumed that the measurements have the characteristics of an equal interval scale. The assumption that Likert-type scales are equal interval scales in statistical analyses is a widely accepted and theoretically valid approach in the social sciences (Carifio & Perla, 2008).
- The assumption of homogeneity of variance has been met. Homogeneity of variance, one of the basic assumptions of parametric tests (t-test and ANOVA) used in comparisons of groups, has been taken into account. Homogeneity of variance means that the variances of the groups being compared are equal or similar to each other, and it increases the reliability of the analysis results (Çelikten Demirel, et al., 2025). It is assumed that sampling errors are at a controllable level. Samples selected using the random sampling method have higher statistical power in terms of representing the population (Büyüköztürk, 2022).
- It is assumed that the data consist of independent observations. In this context, it is accepted that each participant responds to the survey only once and that the observations are completely independent of each other. The assumption of independent observations is a fundamental condition for the valid application of parametric statistical analyses such as t-tests and ANOVA, and a violation of this assumption can negatively affect the reliability and interpretability of the analysis results (Field, 2018).

Limitations

- This research has several methodological and practical limitations in terms of research design, sample structure, and data collection process.
- The research is limited to students studying at Bilecik Şeyh Edebali University between December 1st and 31st, 2025.
- It is not possible to generalize the findings to different universities, faculties, or student groups with different socio-cultural characteristics.
- The fact that the data collection process in this study was carried out within a specific and relatively short time frame may limit the long-term validity of the findings due to the dynamic and rapidly changing nature of social media usage habits.

- The limited sample size of 388 students suggests that the statistical power may be somewhat lower compared to studies conducted with larger samples (Keskin, 2020).
- The study only considered the Instagram platform, thus limiting the representativeness of the results in terms of digital media usage behaviors as a whole.
- The data was collected based on self-report. This may affect the results.

Definitions

Social Media: Social media refers to platforms that enable individuals to interact, share content, and communicate in a digital environment (Okmeydan, 2020).

Instagram: With its visually focused structure, it allows users to share their experiences and gain motivation by interacting with others (Aktan, 2018).

Seeking Gratification: This is a concept encompassing the motives and motivations that drive individuals to use media. In this context, various needs such as social interaction, archiving, self-expression, escapism, and surveillance shape the relationships and usage behaviors of users with media (Saatçioğlu & İnanç, 2020).

Uses and Gratifications Theory: This approach views media users not as passive recipients, but as conscious, active, and purposeful individuals in their media use. This theory suggests that users choose media to satisfy specific needs and exhibit purposeful behaviors in their consumption processes (DeFleur, 2016).

The methodology of this study, the selected sample, and the independent and dependent variables determined by the applied data collection process allow for a systematic, consistent, and reliable examination of the effects of Instagram users' gratification-seeking on the frequency of their platform use. Analyses conducted within the framework of Uses and Gratifications Theory reveal users' fundamental motivations for social media use and provide a more holistic perspective on Instagram usage behavior. Considering the limitations and assumptions of the research, it is assessed that the findings are valid within the context of the Instagram platform; however, comparative studies can be conducted on different social media platforms using similar theoretical frameworks and methods. In this respect, the research contributes to the literature and provides a guiding theoretical and methodological foundation for future social media-focused studies.

FINDINGS

A total of 417 students were reached through an online survey as part of the research. However, when checking the suitability of the dataset for analysis, 29 survey forms containing missing data were excluded from the evaluation. After the data cleaning process, the remaining data from 388 participants (100%), all of whom were students of Bilecik Şeyh Edebali University, were included in the statistical analyses.

Table 1: Demographic Information Distribution

Variables	Groups	Number (n)	Percentage (%)
Gender	Female	231	59.5
	Male	157	40.5
Age Range	18-25	364	93.8
	26-34	11	2.8
	Other	13	3.4
Educational Status	High School and Below	9	2.3
	Associate Degree	74	19.1
	Bachelor's	291	75.0
	Postgraduate	14	3.6
Daily Instagram Usage Period	Less than 30 min	23	5.9
	30 min- 1 hour	60	15.5

	1- 2 hours	116	29.9
	2- 3 hours	98	25.3
	More than 3 hours	91	23.5
Instagram Experience	Less than 1 year	13	3.4
	1-3 years	38	9.8
	3-5 years	110	28.4
	5-7 years	127	32.7
	More than 7 years	100	25.8
Total		388	100.0

Source: Created by the author.

Table 1, which shows the demographic distribution, reveals that a total of 388 people participated in the study, with 59.5% being female and 40.5% male. When evaluating the age distribution of the participants, it is seen that the vast majority of the sample (93.8%) is in the 18-25 age range; followed by other age groups with 3.4% and the 26-34 age range with 2.8%. Regarding educational levels, it was determined that 75% of the participants had a bachelor's degree, 19.1% had an associate's degree, and 3.6% had a postgraduate degree. Regarding Instagram usage frequency, 29.9% of participants use the platform for 1-2 hours a day, 25.3% for 2-3 hours, and 23.5% for more than 3 hours. In terms of experience, 32.7% of participants have been Instagram users for 5-7 years, and 28.4% have been users for 3-5 years.

Descriptive Statistics Regarding the Demographic Characteristics of Participants

The frequency and percentage distribution of participants' demographic characteristics are presented in the tables below.

Table 2: Gender Frequency and Percentage Distribution

Gender	**Frequency (n)**	**Percentage (%)**
Female	231	59,5
Male	157	40,5
Total	388	100,0

Source: Created by the author.

Looking at Table 2 regarding gender, it can be seen that 231 (59.5%) of the participants were women and 157 (40.5%) were men. Accordingly, it is understood that the majority of the individuals participating in the study were women.

Table 3: Age Range Frequency and Percentage Distribution

Age Range	**Frequency (n)**	**Percentage (%)**
18–25	364	93,8
26–34	11	2,8
Other	13	3,4
Total	388	100,0

Source: Created by the author.

Looking at Table 3 regarding the age ranges of the participants, it can be seen that 364 people (93.8%) are in the 18–25 age range. In contrast, 11 participants (2.8%) are in the 26–34 age range, and 13 participants (3.4%) are in other age groups. These findings indicate that the sample is largely composed of people aged 18–25.

Table 4: Frequency and Percentage Distribution of Educational Attainment

Educational Status	Frequency (n)	Percentage (%)
High School and Below	9	2,3
Associate Degree	74	19,1
Bachelor's	291	75,0
Postgraduate	14	3,6
Total	388	100,0

Source: Created by the author.

Looking at Table 4 regarding educational status, it can be seen that 291 participants (75.0%) have a bachelor's degree. This is followed by 74 participants (19.1%) with an associate's degree. The number of participants with postgraduate education is 14 (3.6%), while the number of participants with a high school education or lower is limited to 9 (2.3%). The findings reveal that the sample largely consists of individuals with a bachelor's degree.

Table 5: Daily Instagram Usage Time, Frequency, and Percentage Distribution

Daily Instagram Usage Period	Frequency (n)	Percentage (%)
Less than 30 min	23	5,9
30 min- 1 hour	60	15,5
1- 2 hours	116	29,9
2- 3 hours	98	25,3
More than 3 hours	91	23,5
Total	388	100,0

Source: Created by the author.

Looking at Table 5 regarding participants' daily Instagram usage time, it can be seen that 116 people (29.9%) use Instagram for 1-2 hours a day. This is followed by 98 participants (25.3%) using it for 2-3 hours and 91 participants (23.5%) using it for more than 3 hours. Regarding shorter usage times, 60 participants (15.5%) stated that they use Instagram for 30 minutes to 1 hour, while 23 participants (5.9%) stated that they use it for less than 30 minutes. Overall, it can be said that a significant portion of the participants actively use Instagram for at least 1 hour or more per day.

Table 6: Instagram Experience and Percentage Distribution

Instagram Experience	Frequency (n)	Percentage (%)
Less than 1 year	13	3,4
1-3 years	38	9,8
3-5 years	110	28,4
5-7 years	127	32,7
More than 7 years	100	25,8
Total	388	100,0

Source: Created by the author.

Looking at Table 6 regarding Instagram experience, it can be seen that 127 participants (32.7%) have been using Instagram for 5–7 years. This is followed by 110 participants (28.4%) with 3–5 years and 100 participants (25.8%) with more than 7 years of use. When examining shorter-term users, it is understood that 38 participants (9.8%) have been using Instagram for 1–3 years and 13 participants (3.4%) have been using it for less than 1 year. These findings indicate that a large proportion of participants have been actively using Instagram for a long time.

Descriptive Statistics Regarding Scale Items

The frequency and percentage distribution of participants' levels of agreement with the statements included in the research scale are presented in the tables below.

Table 7: "To interact with some people" Frequency and Percentage Distribution

Response Options	Frequency (n)	Percentage (%)
Strongly Disagree	44	11,3
Disagree	45	11,6

Undecided	40	10,3
Agree	191	49,2
Strongly Agree	68	17,5
Total	388	100,0

Source: Created by the author.

Looking at Table 7 regarding interacting with certain individuals, it is seen that 191 participants (49.2%) responded "agree" and 68 (17.5%) responded "strongly agree". Accordingly, a total of 259 participants (66.7%) positively agreed with this statement. In contrast, 45 participants (11.6%) responded "disagree" and 44 participants (11.3%) responded "strongly disagree". This indicates that a total of 89 participants (22.9%) did not agree with the statement. Furthermore, 40 participants (10.3%) stated that they were undecided about this statement. Overall, the findings suggest that a large proportion of participants conducted this activity with the aim of interacting with certain individuals.

Table 8: "To maintain good relationships with others" Frequency and Percentage Distribution

Response Options	Frequency (n)	Percentage (%)
Strongly Disagree	55	14,2
Disagree	104	26,8
Undecided	75	19,3
Agree	114	29,4
Strongly Agree	40	10,3
Total	388	100,0

Source: Created by the author.

According to the findings presented in Table 8, 114 participants (29.4%) responded "agree," and 40 (10.3%) responded "strongly agree." Thus, a total of 154 participants (39.7%) positively agreed with the statement. In contrast, 104 participants (26.8%) responded "disagree," and 55 participants (14.2%) responded "strongly disagree." This indicates that a total of 159 participants (41.0%) disagreed with the statement. Furthermore, 75 participants (19.3%) stated they were undecided. Overall, the findings suggest that a large majority of participants undertook this action with the aim of maintaining good relationships with others.

Table 9: "To receive updates about my family and close friends" Frequency and Percentage Distribution

Response Options	Frequency (n)	Percentage (%)
Strongly Disagree	57	14,7
Disagree	50	12,9
Undecided	32	8,2
Agree	177	45,6
Strongly Agree	72	18,6
Total	388	100,0

Source: Created by the author.

Looking at Table 9 regarding receiving updates about family and close friends, 177 participants (45.6%) responded "agree," and 72 (18.6%) responded "strongly agree." Therefore, a total of 249 participants (64.2%) positively agreed with this statement. In contrast, 50 participants (12.9%) responded "disagree," and 57 participants (14.7%) responded "strongly disagree." This indicates that a total of 107 participants (27.6%) disagreed with the statement. Furthermore, 32 participants (8.2%) stated they were undecided. Overall, the findings reveal that a significant portion of participants use the platform to follow updates about their family and close friends.

Table 10: "To stay connected with my family and friends" Frequency and Percentage Distribution

Response Options	Frequency (n)	Percentage (%)
Strongly Disagree	57	14,7
Disagree	33	8,5
Undecided	33	8,5
Agree	191	49,2

Strongly Agree	74	19,1
Total	388	100,0

Source: Created by the author.

Looking at Table 10 regarding staying connected with family and friends, 191 participants (49.2%) responded "agree," and 74 (19.1%) responded "strongly agree." Therefore, a total of 265 participants (68.3%) positively agreed with this statement. In contrast, 33 participants (8.5%) responded "disagree," and 57 participants (14.7%) responded "strongly disagree." This indicates that a total of 90 participants (23.2%) disagreed with the statement. Furthermore, 33 participants (8.5%) stated they were undecided. Overall, the findings reveal that the majority of participants use the platform to stay connected with family and friends.

Table 11: "To communicate with my family and friends."
Frequency and Percentage Distribution

Response Options	Frequency (n)	Percentage (%)
Strongly Disagree	54	13,9
Disagree	41	10,6
Undecided	34	8,8
Agree	180	46,4
Strongly Agree	79	20,4
Total	388	100,0

Source: Created by the author.

Looking at Table 11 regarding communicating with family and friends, it can be seen that 180 participants (46.4%) responded "agree" and 79 (20.4%) responded "strongly agree". Accordingly, a total of 259 participants (66.8%) positively agreed with this statement. In contrast, 41 participants (10.6%) responded "disagree" and 54 participants (13.9%) responded "strongly disagree". This indicates that a total of 95 participants (24.5%) disagreed with the statement. Furthermore, 34 participants (8.8%) stated that they were undecided about this statement. Overall, the findings show that a large majority of participants use the platform to communicate with family and friends.

Table 12: "Because the people around me use Instagram."
Frequency and Percentage Distribution

Response Options	Frequency (n)	Percentage (%)
Strongly Disagree	105	27,1
Disagree	101	26,0
Undecided	49	12,6
Agree	98	25,3
Strongly Agree	35	9,0
Total	388	100,0

Source: Created by the author.

Looking at Table 12, which is related to the fact that people around me use Instagram, it is seen that 98 participants (25.3%) responded "agree" and 35 (9.0%) responded "strongly agree". Accordingly, a total of 133 participants (34.3%) showed positive agreement with the statement. In contrast, 101 participants (26.0%) responded "disagree" and 105 participants (27.1%) responded "strongly disagree". This shows that a total of 206 participants (53.1%) did not agree with the statement. In addition, 49 participants (12.6%) stated that they were undecided about this statement. Overall, the findings indicate that a large majority of participants do not see Instagram use as a motivation based on social network influence.

Table 13: "To be aware of what's going on around me"
Frequency and Percentage Distribution

Response Options	Frequency (n)	Percentage (%)
Strongly Disagree	42	10,8
Disagree	29	7,5
Undecided	30	7,7

Agree	173	44,6
Strongly Agree	114	29,4
Total	388	100,0

Source: Created by the author.

Looking at Table 13 regarding staying informed about what's happening around them, 173 participants (44.6%) responded "agree," and 114 (29.4%) responded "strongly agree." Therefore, a total of 287 participants (74.0%) positively agreed with this statement. In contrast, 29 participants (7.5%) responded "disagree," and 42 participants (10.8%) responded "strongly disagree." This indicates that 71 participants (18.3%) disagreed with the statement. Furthermore, 30 participants (7.7%) stated they were undecided. Overall, the findings show that a large majority of participants use Instagram to stay informed about what's happening around them.

Table 14: "To connect/meet people who share similar interests with me." Frequency and Percentage Distribution

Response Options	Frequency (n)	Percentage (%)
Strongly Disagree	76	19,6
Disagree	81	20,9
Undecided	53	13,7
Agree	125	32,2
Strongly Agree	53	13,7
Total	388	100,0

Source: Created by the author.

Looking at Table 14 regarding connecting/meeting people with similar interests, 125 participants (32.2%) responded "agree," and 53 (13.7%) responded "strongly agree." Therefore, a total of 178 participants (45.9%) positively agreed with this statement. In contrast, 81 participants (20.9%) responded "disagree," and 76 participants (19.6%) responded "strongly disagree." This indicates that a total of 157 participants (40.5%) disagreed with the statement. Furthermore, 53 participants (13.7%) stated they were undecided. Overall, the findings show that a large majority of participants used this method to connect with and meet people who share similar interests.

Table 15: "To record everyday events through photographs." Frequency and Percentage Distribution

Response Options	Frequency (n)	Percentage (%)
Strongly Disagree	69	17,8
Disagree	65	16,8
Undecided	52	13,4
Agree	143	36,9
Strongly Agree	59	15,2
Total	388	100,0

Source: Created by the author.

Looking at Table 15 regarding recording everyday events through photographs, it is seen that 143 participants (36.9%) responded "agree" and 59 (15.2%) responded "strongly agree". Accordingly, a total of 202 participants (52.1%) positively agreed with this statement. In contrast, 65 participants (16.8%) responded "disagree" and 69 participants (17.8%) responded "strongly disagree". This indicates that a total of 134 participants (34.6%) did not agree with the statement. Furthermore, 52 participants (13.4%) stated that they were undecided on this statement. Overall, the findings show that a large proportion of participants use Instagram to record everyday events through photographs.

Table 16: "To create my personal space" Frequency and Percentage Distribution

Response Options	Frequency (n)	Percentage (%)
Strongly Disagree	62	16,0
Disagree	71	18,3
Undecided	78	20,1
Agree	120	30,9
Strongly Agree	57	14,7

Total	388	100,0

Source: Created by the author.

Looking at Table 16 regarding creating personal space, it can be seen that 120 participants (30.9%) responded "agree" and 57 (14.7%) responded "strongly agree". Accordingly, a total of 177 participants (45.6%) positively agreed with this statement. In contrast, 71 participants (18.3%) responded "disagree" and 62 participants (16.0%) responded "strongly disagree". This indicates that a total of 133 participants (34.3%) disagreed with the statement. Furthermore, 78 participants (20.1%) stated that they were undecided on this statement. Overall, the findings suggest that a large majority of participants use Instagram to create a personal space.

Table 17: "To record my tracks with a map made of photographs." Frequency and Percentage Distribution

Response Options	Frequency (n)	Percentage (%)
Strongly Disagree	70	18,0
Disagree	80	20,6
Undecided	62	16,0
Agree	122	31,4
Strongly Agree	54	13,9
Total	388	100,0

Source: Created by the author.

Looking at Table 17 regarding recording their tracks using a map of photographs, it is seen that 122 participants (31.4%) responded "agree" and 54 (13.9%) responded "strongly agree". Accordingly, a total of 176 participants (45.3%) positively agreed with this statement. In contrast, 80 participants (20.6%) responded "disagree" and 70 participants (18.0%) responded "strongly disagree". This indicates that a total of 150 participants (38.6%) did not agree with the statement. Furthermore, 62 participants (16.0%) stated that they were undecided on this matter. Overall, the findings show that participants used a map created from photographs to record their tracks.

Table 18: "To start a personal blog"
Frequency and Percentage Distribution

Response Options	Frequency (n)	Percentage (%)
Strongly Disagree	86	22,2
Disagree	95	24,5
Undecided	72	18,6
Agree	94	24,2
Strongly Agree	41	10,6
Total	388	100,0

Source: Created by the author.

Looking at Table 18 regarding personal blogging, it is seen that 94 participants (24.2%) responded "agree" and 41 (10.6%) responded "strongly agree". Accordingly, a total of 135 participants (34.8%) positively agreed with this statement. In contrast, 95 participants (24.5%) responded "disagree" and 86 participants (22.2%) responded "strongly disagree". This indicates that a total of 181 participants (46.7%) disagreed with the statement. Furthermore, 72 participants (18.6%) stated that they were undecided. Overall, the findings reveal that participants have a low tendency to use blogging for personal blogging purposes.

Table 19: "To take fancy photos and save them online"
Frequency and Percentage Distribution

Response Options	Frequency (n)	Percentage (%)
Strongly Disagree	103	26,5
Disagree	100	25,8
Undecided	57	14,7
Agree	87	22,4
Strongly Agree	41	10,6
Total	388	100,0

Source: Created by the author.

Looking at Table 19 regarding taking fancy photos and saving them online, 87 participants (22.4%) responded "agree," and 41 (10.6%) responded "strongly agree." Therefore, a total of 128 participants (33.0%) positively agreed with the statement. In contrast, 100 participants (25.8%) responded "disagree," and 103 participants

(26.5%) responded "strongly disagree." This indicates that a total of 203 participants (52.3%) disagreed with the statement. Furthermore, 57 participants (14.7%) stated they were undecided. Overall, the findings suggest that a large proportion of participants have a low tendency to take fancy photos and use them for online saving.

Table 20: "To upload photos and videos with various filters applied." Frequency and Percentage Distribution

Response Options	Frequency (n)	Percentage (%)
Strongly Disagree	110	28,4
Disagree	112	28,9
Undecided	59	15,2
Agree	78	20,1
Strongly Agree	29	7,5
Total	388	100,0

Source: Created by the author.

Looking at Table 20 regarding uploading photos and videos with various filters applied, it is seen that 78 participants (20.1%) responded "agree" and 29 (7.5%) responded "strongly agree". Accordingly, a total of 107 participants (27.6%) positively agreed with this statement. In contrast, 112 participants (28.9%) responded "disagree" and 110 participants (28.4%) responded "strongly disagree". This shows that a total of 222 participants (57.3%) did not agree with the statement. In addition, 59 participants (15.2%) stated that they were undecided about this statement. Overall, the findings reveal that the majority of participants have a low tendency to use the platform for uploading filtered photos and videos.

Table 21: "To enable my update" Frequency and Percentage Distribution

Response Options	Frequency (n)	Percentage (%)
Strongly Disagree	81	20,9
Disagree	84	21,6
Undecided	72	18,6
Agree	114	29,4
Strongly Agree	37	9,5
Total	388	100,0

Source: Created by the author.

Looking at Table 21 regarding staying up-to-date, it can be seen that 114 participants (29.4%) responded "agree" and 37 (9.5%) responded "strongly agree". Accordingly, a total of 151 participants (38.9%) positively agreed with this statement. In contrast, 84 participants (21.6%) responded "disagree" and 81 participants (20.9%) responded "strongly disagree". This indicates that a total of 165 participants (42.5%) did not agree with the statement. Furthermore, 72 participants (18.6%) stated that they were undecided about this statement. Overall, the findings reveal that participants have a moderate tendency to use this method for staying up-to-date.

Table 22: "To be noticed by others" Frequency and Percentage Distribution

Response Options	Frequency (n)	Percentage (%)
Strongly Disagree	149	38,4
Disagree	137	35,3
Undecided	55	14,2
Agree	31	8,0
Strongly Agree	16	4,1
Total	388	100,0

Source: Created by the author.

Looking at Table 22 regarding being noticed by others, 31 participants (8.0%) responded "agree," and 16 (4.1%) responded "strongly agree." Therefore, a total of 47 participants (12.1%) positively agreed with this statement. In contrast, 137 participants (35.3%) responded "disagree," and 149 participants (38.4%) responded "strongly disagree." This indicates that a total of 286 participants (73.7%) disagreed with the statement. Furthermore, 55 participants (14.2%) stated they were undecided. Overall, the findings reveal that a large proportion of participants have a very low tendency to use Instagram for the purpose of being noticed by others.

Table 23: "To express my true self (who I really am)" Frequency and Percentage Distribution

Response Options	Frequency (n)	Percentage (%)
Strongly Disagree	126	32,5
Disagree	127	32,7
Undecided	63	16,2
Agree	48	12,4
Strongly Agree	24	6,2
Total	388	100,0

Source: Created by the author.

Looking at Table 23 regarding expressing my true self (who I really am), 48 participants (12.4%) responded "agree," and 24 (6.2%) responded "strongly agree." Therefore, a total of 72 participants (18.6%) positively agreed with this statement. In contrast, 127 participants (32.7%) responded "disagree," and 126 participants (32.5%) responded "strongly disagree." This indicates that a total of 253 participants (65.2%) disagreed with the statement. Furthermore, 63 participants (16.2%) stated they were undecided. Overall, the findings reveal that a large proportion of participants have a low tendency to use Instagram to express their true selves.

Table 24: "To share my personal information with others" Frequency and Percentage Distribution

Response Options	Frequency (n)	Percentage (%)
Strongly Disagree	158	40,7
Disagree	138	35,6
Undecided	49	12,6
Agree	34	8,8
Strongly Agree	9	2,3
Total	388	100,0

Source: Created by the author.

Looking at Table 24 regarding sharing personal information with others, 34 participants (8.8%) responded "agree," and 9 (2.3%) responded "strongly agree." Therefore, a total of 43 participants (11.1%) positively agreed with this statement. In contrast, 138 participants (35.6%) responded "disagree," and 158 participants (40.7%) responded "strongly disagree." This indicates that a total of 296 participants (76.3%) disagreed with the statement. Furthermore, 49 participants (12.6%) stated they were undecided. Overall, the findings reveal that a large proportion of participants have a very low tendency to use Instagram to share their personal information with others.

Table 25: "To reveal what I have" Frequency and Percentage Distribution

Response Options	Frequency (n)	Percentage (%)
Strongly Disagree	164	42,3
Disagree	135	34,8
Undecided	47	12,1
Agree	32	8,2
Strongly Agree	10	2,6
Total	388	100,0

Source: Created by the author.

Looking at Table 25 regarding showing what they own, 32 participants (8.2%) responded "agree," and 10 (2.6%) responded "strongly agree." Therefore, a total of 42 participants (10.8%) positively agreed with this statement. In contrast, 135 participants (34.8%) responded "disagree," and 164 participants (42.3%) responded "strongly disagree." This indicates that a total of 299 participants (77.1%) disagreed with the statement. Furthermore, 47 participants (12.1%) stated they were undecided. Overall, the findings suggest that the tendency of a large proportion of participants to use Instagram to show what they own is very low.

Table 26: "To escape reality" Frequency and Percentage Distribution

Response Options	Frequency (n)	Percentage (%)
Strongly Disagree	147	37,9
Disagree	130	33,5
Undecided	48	12,4

Agree	39	10,1
Strongly Agree	24	6,2
Total	388	100,0

Source: Created by the author.

Looking at Table 26 regarding escaping reality, it is seen that 39 participants (10.1%) responded "agree" and 24 (6.2%) responded "strongly agree". Accordingly, a total of 63 participants (16.3%) positively agreed with the statement. In contrast, 130 participants (33.5%) responded "disagree" and 147 participants (37.9%) responded "strongly disagree". This shows that a total of 277 participants (71.4%) did not agree with the statement. In addition, 48 participants (12.4%) stated that they were undecided about this statement. Overall, the findings indicate that the tendency of a large proportion of participants to use this as a means of escaping reality is at a very low level.

Table 27: "To forget my problems" Frequency and Percentage Distribution

Response Options	Frequency (n)	Percentage (%)
Strongly Disagree	112	28,9
Disagree	99	25,5
Undecided	69	17,8
Agree	68	17,5
Strongly Agree	40	10,3
Total	388	100,0

Source: Created by the author.

Looking at Table 27 regarding forgetting problems, it is seen that 68 participants (17.5%) responded "agree" and 40 (10.3%) responded "strongly agree". Accordingly, a total of 108 participants (27.8%) positively agreed with this statement. In contrast, 99 participants (25.5%) responded "disagree" and 112 participants (28.9%) responded "strongly disagree". This shows that a total of 211 participants (54.4%) did not agree with the statement. Furthermore, 69 participants (17.8%) stated that they were undecided about this statement. Overall, it can be said that more than half of the participants do not see Instagram use as a tool for forgetting their problems, but a significant portion do use the platform for this purpose.

Table 28: "To escape loneliness" Frequency and Percentage Distribution

Response Options	Frequency (n)	Percentage (%)
Strongly Disagree	136	35,1
Disagree	121	31,2
Undecided	68	17,5
Agree	41	10,6
Strongly Agree	22	5,7
Total	388	100,0

Source: Created by the author.

Looking at Table 28 regarding overcoming loneliness, it is seen that 41 participants (10.6%) responded "agree" and 22 (5.7%) responded "strongly agree". Accordingly, a total of 63 participants (16.3%) positively agreed with this statement. In contrast, 121 participants (31.2%) responded "disagree" and 136 participants (35.1%) responded "strongly disagree". This indicates that a total of 257 participants (66.3%) did not agree with the statement. Furthermore, 68 participants (17.5%) stated that they were undecided on this statement. Overall, it can be said that a large portion of the participants do not use Instagram to overcome loneliness, but a small segment does use the platform for this purpose.

Table 29: "To achieve what I want without putting in too much effort" Frequency and Percentage Distribution

Response Options	Frequency (n)	Percentage (%)
Strongly Disagree	136	35,1
Disagree	128	33,0
Undecided	53	13,7
Agree	50	12,9
Strongly Agree	21	5,4
Total	388	100,0

Source: Created by the author.

Looking at Table 29 regarding achieving what one wants without much effort, 50 participants (12.9%) responded "agree," and 21 (5.4%) responded "strongly agree." Therefore, a total of 71 participants (18.3%) positively agreed with this statement. In contrast, 128 participants (33.0%) responded "disagree," and 136 participants (35.1%) responded "strongly disagree." This indicates that a total of 264 participants (68.1%) disagreed with the statement. Furthermore, 53 participants (13.7%) stated they were undecided. Overall, this suggests that a large majority of participants do not view Instagram use as a means of achieving goals without much effort.

Table 30: "To relax" Frequency and Percentage Distribution

Response Options	Frequency (n)	Percentage (%)
Strongly Disagree	68	17,5
Disagree	59	15,2
Undecided	63	16,2
Agree	141	36,3
Strongly Agree	57	14,7
Total	388	100,0

Source: Created by the author.

Looking at Table 30 regarding relaxation, it is seen that 141 participants (36.3%) responded "agree" and 57 (14.7%) responded "strongly agree". Accordingly, a total of 198 participants (51.0%) positively agreed with this statement. In contrast, 59 participants (15.2%) responded "disagree" and 68 participants (17.5%) responded "strongly disagree". This indicates that a total of 127 participants (32.7%) did not agree with the statement. Furthermore, 63 participants (16.2%) stated that they were undecided on this statement. Overall, this suggests that a large majority of participants use Instagram for relaxation purposes.

Table 31: "To browse photos related to my interests."
Frequency and Percentage Distribution

Response Options	Frequency (n)	Percentage (%)
Strongly Disagree	66	17,0
Disagree	56	14,4
Undecided	49	12,6
Agree	142	36,6
Strongly Agree	75	19,3
Total	388	100,0

Source: Created by the author.

Looking at Table 31 regarding browsing photos related to their interests, it is seen that 142 participants (36.6%) responded "agree" and 75 (19.3%) responded "strongly agree". Accordingly, a total of 217 participants (55.9%) positively agreed with this statement. In contrast, 56 participants (14.4%) responded "disagree" and 66 participants (17.0%) responded "strongly disagree". Thus, it is understood that a total of 122 participants (31.4%) did not agree with the statement. Furthermore, 49 participants (12.6%) stated that they were undecided on this statement. Overall, this reveals that a large majority of participants prefer using Instagram to follow visual content related to their interests.

Table 32: "To get a glimpse into the daily lives of celebrities."
Frequency and Percentage Distribution

Response Options	Frequency (n)	Percentage (%)
Strongly Disagree	130	33,5
Disagree	94	24,2
Undecided	64	16,5
Agree	71	18,3
Strongly Agree	29	7,5
Total	388	100,0

Source: Created by the author.

Looking at Table 32 regarding glimpsing the daily lives of celebrities, it is seen that 71 participants (18.3%) responded "agree" and 29 (7.5%) responded "strongly agree". Accordingly, a total of 100 participants (25.8%) positively agreed with this statement. In contrast, 94 participants (24.2%) responded "disagree" and 130 participants (33.5%) responded "strongly disagree". Therefore, it is understood that a total of 224 participants (57.7%) disagreed with the statement. Furthermore, 64 participants (16.5%) stated that they were undecided on

this statement. Overall, this indicates that a large majority of participants do not view Instagram use as a means of glimpsing the daily lives of celebrities.

Table 33: "To browse different, ornate photos" Frequency and Percentage Distribution

Response Options	Frequency (n)	Percentage (%)
Strongly Disagree	109	28,1
Disagree	90	23,2
Undecided	48	12,4
Agree	103	26,5
Strongly Agree	38	9,8
Total	388	100,0

Source: Created by the author.

Looking at Table 33 regarding browsing different, ornate photos, it is seen that 103 participants (26.5%) responded "agree" and 38 (9.8%) responded "strongly agree". Accordingly, a total of 141 participants (36.3%) positively agreed with this statement. In contrast, 90 participants (23.2%) responded "disagree" and 109 participants (28.1%) responded "strongly disagree". Thus, it is understood that a total of 199 participants (51.3%) disagreed with the statement. Furthermore, 48 participants (12.4%) stated that they were undecided on this statement. Overall, this indicates that a large majority of participants do not use Instagram for the purpose of browsing different, ornate photos.

Table 34: "To get a glimpse into the daily lives of people around the world." Frequency and Percentage Distribution

Response Options	Frequency (n)	Percentage (%)
Strongly Disagree	73	18,8
Disagree	75	19,3
Undecided	56	14,4
Agree	122	31,4
Strongly Agree	62	16,0
Total	388	100,0

Source: Created by the author.

Looking at Table 34 regarding browsing the daily lives of people around the world, 122 participants (31.4%) responded "agree," and 62 (16.0%) responded "strongly agree." Therefore, a total of 184 participants (47.4%) positively agreed with the statement. In contrast, 75 participants (19.3%) responded "disagree," and 73 participants (18.8%) responded "strongly disagree." Thus, a total of 148 participants (38.1%) disagreed with the statement. Furthermore, 56 participants (14.4%) indicated they were undecided. Overall, this suggests that a large majority of participants use Instagram to browse the daily lives of people around the world.

Average of the items

The study examined whether the dataset exhibited a normal distribution, and the descriptive statistics and normality test results for the scale sub-dimensions are presented below.

Table 35: Descriptive Statistics and Normality Test Results for the Scale Sub-Dimensions

Sub-dimensions	N	Mean	Std. Deviation	Skewness	Kurtosis
Social Interaction	388	3.28	0.99	-0.64	-0.26
Archiving	388	2.86	1.10	-0.09	-0.91
Surveillance	388	2.86	1.13	0.06	-1.01
Escape	388	2.45	1.00	0.48	-0.42
Self-Expression	388	2.21	0.93	0.79	0.31

Source: Created by the author.

Table 35, which shows the descriptive statistics and normality test results for the sub-dimensions of the scale, indicates that the skewness and kurtosis values for the sub-dimensions of the scale used in the study range from +1.5 to -1.5. These values show that the assumption of normal distribution of the data is met. Therefore, parametric

tests (T-test and ANOVA) were used in hypothesis testing. When the participants' search for satisfaction was examined, it was observed that the highest average was in the Social Interaction dimension ($\bar{x}$=3.28), followed by Archiving ($\bar{x}$=2.86) and Surveillance ($\bar{x}$=2.86) dimensions, respectively. The lowest search for satisfaction among the participants was found in the Self-Expression dimension ($\bar{x}$=2.21).

Difference Analyses Based on Demographic Variables

In this study, T-tests and ANOVA analyses were used to examine whether there was a statistically significant difference between the demographic characteristics (gender and education level) of the participants and their satisfaction levels. The findings are presented below.

Table 36: T-test by Gender

Alt Boyutlar	Grnder	N	Mean	Std. Deviation	t	p (Sig.)
Social Interaction	Female	231	3.34	0.95	-1.60	.110
	Male	157	3.18	1.05		
Archiving	Female	231	3.08	1.07	-4.77	<.001
	Male	157	2.55	1.07		
Self-Expression	Female	231	2.27	0.91	-1.35	.179
	Male	157	2.14	0.95		
Escape	Female	231	2.45	0.97	-0.08	.937
	Male	157	2.44	1.04		
Surveillance	Female	231	3.00	1.10	-3.15	.002
	Male	157	2.64	1.15		

Source: Created by the author.

When examining Table 36, which shows the results of the T-test by gender, a statistically significant difference was found between the gender of the participants and their gratification seeking in Archiving (t=-4.77, p<.05) and Surveillance (t=-3.15, p<.05). When the means are examined, the Archiving gratification score of female participants ($\bar{x}$=3.08) is significantly higher than that of male participants ($\bar{x}$=2.55).

Similarly, the Surveillance gratification score of females ($\bar{x}$=3.00) is higher than that of males ($\bar{x}$=2.64). On the other hand, no statistically significant difference was found between genders in the sub-dimensions of Social Interaction, Self-Expression, and Escape (p>.05). This indicates that these gratification seeking patterns are similar regardless of gender.

Table 37: Eğitim Durumuna Göre ANOVA Sonuçları

Independent Variable	Sub-dimensions	Sum of Squares	Mean Squares	F	p (Sig.)	Tukey
Educational Status	Social Interaction	4.891	1.630	0.967	**.408**	No difference
Educational Status	Archiving	4.449	1.483	0.837	**.474**	No difference
Educational Status	Self-Expression	4.270	1.423	1.080	**.358**	No difference
Educational Status	Escape	6.851	2.284	1.450	**.228**	No difference
Educational Status	Surveillance	4.030	1.343	0.721	**.540**	No difference

Source: Created by the author.

When examining Table 37, which shows the ANOVA results according to educational status, no statistically significant difference was found in any sub-dimension of Instagram gratification seeking (Social Interaction,

Archiving, Self-Expression, Escape, and Surveillance) according to the Educational Status variable ($p > .05$).This indicates that even if the educational levels (Associate Degree, Bachelor's Degree, or Postgraduate Degree) of university students vary, their motivations for using Instagram are similar.

Findings Related to Hypothesis 1

The first hypothesis of the study is: "Is there a significant difference between social interaction satisfaction and the frequency of Instagram use?" Within this scope, the findings were examined for each sub-dimension. Participant opinions regarding the first hypothesis are presented in Table 38.

Table 38: ANOVA Results of Items Constituting the Social Interaction Dimension Based on Instagram Usage Frequency

Social Interaction Items	Sum of Squares	df	Mean Squares	F	p (Sig.)	Tukey
To interact with specific individuals	8.82	4	2.20	1.46	.214	No difference
To maintain positive relationships with others	23.24	4	5.81	3.88	.004	3h+> 30min-1h
To get updates about my family and friends	37.47	4	9.37	5.59	.000	3h+> 30min-
To stay connected with my family and friends	31.56	4	7.89	4.86	.000	3h +> 30min-
To communicate with my family and friends	28.15	4	7.04	4.26	.002	3h +> 30min-
Because people around me use Instagram	25.44	4	6.36	3.58	.007	3ht+> 1-2 hours
To stay informed about what's going on around me	20.52	4	5.13	3.32	.011	3h +> 30min-
To connect/meet people with similar interests	30.47	4	7.62	4.22	.002	3h +> 30min-1h

Source: Created by the author.

When examining Table 38, which shows the ANOVA results of the items constituting the social interaction dimension according to Instagram usage frequency, statistically significant differences were observed in 7 out of 8 items constituting the social interaction dimension according to Instagram usage frequency ($p<0.05$). Only in the item "To interact with some people" was no significant difference found between the groups ($p>.05$).

When the items with significant differences were examined, it was seen that the general trend was that "the level of participation increased as the usage time increased." In particular, in the items "To receive updates about my family and close friends" ($F=5.59$; $p<.001$) and "To stay connected" ($F=4.86$; $p<.001$), the averages of those who use it for more than 3 hours a day are significantly higher than those who use it less (less than 30 minutes).

This situation shows that users who spend more time on Instagram use the platform more intensively as a tool to maintain their connections with their social circles and to follow current events. Accordingly, Hypothesis 1 is accepted.

Findings Related to Hypothesis 2

The second hypothesis of the study is: Is there a significant difference between archiving satisfaction and the frequency of Instagram use? In this context, the findings were examined for each sub-dimension. Participant opinions regarding the second hypothesis are presented in Table 39.

Table 39: ANOVA Results Based on the Frequency of Use of Expressions Constituting the Archiving Dimension

Archiving Statements	Mean Squares	F	p (Sig.)	Tukey
To record everyday events through photos	11.34	6.52	<.001	3h+> 30min-
To create my personal space	6.72	4.05	.003	3h+> 30min-
To record my tracks with a map of photos (e.g., travel)	9.12	5.28	<.001	3h+> 30min-
To do personal blogging	6.56	3.87	.004	3h+> 30min-
To take fancy photos and save them online	7.95	4.46	.002	3h+> 30min-
To upload photos and videos with various filters applied	6.30	3.88	.004	3h+> 30min-

Source: Created by the author.

When examining Table 39, which shows the ANOVA results according to the frequency of use of the statements constituting the archiving dimension, significant differences were observed in all statements constituting the archiving dimension. In all cases of differentiation, the averages of the group using it for more than 3 hours were significantly higher than those using it for less than 30 minutes. In particular, for the items "To record daily events through photographs" ($F=6.52$; $p<.001$) and "To create my personal space" ($F=5.28$; $p<.001$), the averages of those using it for more than 3 hours a day were significantly higher than those using it less (less than 30 minutes). Accordingly, Hypothesis 2 was accepted.

Findings Related to Hypothesis 3

The third hypothesis of the study, "There is a significant difference between satisfaction with self-expression and frequency of Instagram use," was tested. In this context, the expressions constituting the self-expression dimension were examined according to the participants' daily usage times, and the analysis results are presented in Table 40.

Table 40: ANOVA Results Based on Frequency of Use of Self-Expression Statements

Expressions of Self-Expression	Mean Squares	F	p (Sig.)	Tukey
To keep myself updated	9.26	5.68	<.001	3h+> 30min-1h
To be noticed by others	2.85	2.37	.052	No difference
To express my true self	5.56	3.90	.004	3h+> 30min-1h
To share my personal information with others	4.03	3.79	.005	3h+> 30min-1h
To show what I have	2.07	1.89	.111	No difference

Source: Created by the author.

When examining Table 40, which shows the ANOVA results based on the frequency of use of self-expression statements, significant differences were found in all statements except "To be noticed by others" and "To show what I have" based on the frequency of Instagram use. Significant differences emerged between those who used Instagram for more than 3 hours a day and those who used it for 30 minutes to 1 hour. This finding indicates that the level of participation in some self-expression statements increases as the usage time increases. Therefore, Hypothesis 3 is partially accepted.

Findings Related to Hypothesis 4

The fourth hypothesis of the study, "There is a significant difference between escapism satisfaction and the frequency of Instagram use," was tested. In this context, the expressions constituting the escapism dimension were examined according to the participants' daily usage times, and the analysis results are presented in Table 41.

Table 41: ANOVA Results Based on the Frequency of Use of Escape Phrases

Escape Phrases	Mean Squares	F	p (Sig.)	Tukey
To escape reality	2.59	1.80	.128	No difference
To forget my problems	4.50	2.54	.039	3h+> 30min-1h
To escape loneliness	1.22	0.86	.487	No difference
To achieve what I want without much effort	6.86	4.92	.001	3h+> 30min-1h
To relax	5.86	3.37	.010	3h+> 30min-1h

Source: Created by the author.

When examining Table 41, which shows the ANOVA results based on the frequency of use of escape statements, significant differences were found in the statements "to relax," "to achieve what I want without much effort," and "to forget my problems," which constitute the escape dimension, according to the frequency of Instagram use. This difference was observed not in the escape dimension as a whole, but in some sub-dimensions, depending on the frequency of use. These differences emerged between those who used Instagram for more than 3 hours a day and those who used it for 30 minutes to 1 hour. No significant difference was found in the statements "to escape reality" and "to escape loneliness." Accordingly, Hypothesis 4 is partially accepted.

Findings Related to Hypothesis 5

The fifth hypothesis of the study, "There is a significant difference between surveillance satisfaction and the frequency of Instagram use," was tested. In this context, the statements constituting the surveillance dimension were examined according to the participants' daily usage times, and the analysis results are presented in Table 42.

Table 42: ANOVA Results Based on Frequency of Use of Surveillance Statements

Surveillance Statements	Mean Squares	F	p (Sig.)	Tukey
To browse photos related to my interests	6.32	3.42	.009	3 saat+> 30 dk- 1 saat
To take a look at celebrities' daily lives	7.55	4.52	.001	3 saat+> 1- 2 saat
To browse different, glamorous photos	11.12	6.14	<.001	3 saat+> 30 dk altı
To take a look at people's daily lives from all over the world	9.56	5.25	<.001	3 saat+> 30 dk- 1 saat

Source: Created by the author.

Table 42, showing ANOVA differences based on usage performance in monitoring problems, reveals significant differences in all details of the monitoring dimension. Those who use more than 3 devices per day consistently have significantly stronger motivations for monitoring their sleep compared to the 30-1 hour or less than 30-minute groups. Specifically, the average scores for those who use more than 3 devices per day are significantly higher than those who use less (less than 30 minutes) for the reasons "To browse different, fancy photos" ($F=6.14$; $p<.001$) and "To browse the lives of people all over the world" ($F=5.25$; $p<.001$). Therefore, Hypothesis 5 is accepted.

CONCLUSION AND RECOMMENDATIONS

This study examined the differences in Instagram usage satisfaction among university students based on their frequency of use. The findings indicate that the frequency of Instagram use is a determining variable for some types of satisfaction. The study first examined the demographic characteristics of the participants and the differences in their Instagram usage purposes regarding satisfaction. It was determined that the majority of the sample consisted of young adults, and a significant portion of the participants used Instagram for more than 1 hour a day. This indicates that Instagram holds a central place in the daily lives of university students.

Comparisons based on gender revealed that female participants scored significantly higher than male participants in archiving and surveillance satisfaction. Conversely, no significant gender-based differences were found in social interaction, self-expression, and escapism satisfaction. This finding suggests that some Instagram usage purposes are adopted independently of gender.

Analyses based on Instagram usage frequency showed strong correlations between social interaction satisfaction and usage duration. In the vast majority of items comprising the social interaction dimension, participants who used the platform for more than 3 hours a day had higher levels of engagement. This result indicates that Instagram strengthens the function of maintaining social relationships and staying connected with the environment for heavy users. The findings reveal that individuals who use the platform for more than 3 hours a day use it not just for leisure but 'purposefully' to protect their social capital. Therefore, as usage time increases, the platform ceases to be merely an entertainment tool and transforms into a fundamental communication necessity.

When examining the levels of archiving satisfaction and frequency of use, statistically significant differences were found in all items. In particular, individuals who spend a significant amount of time on the platform actively use Instagram to document their daily lives and accumulate memories. This finding proves that Instagram is not only a platform for instant sharing but also functions as a personal digital memory. The increase in archiving satisfaction along with increased usage time reveals that users position the platform as a memory repository, constructing and immortalizing their life stories in the digital environment.

Findings regarding satisfaction with self-expression showed significant differences in some expressions depending on frequency of use. However, this difference was not observed in all items. Therefore, it can be said that satisfaction with self-expression is only partially related to the duration of Instagram use. This suggests that motivation for self-expression may depend on more individual and contextual factors among users.

In terms of escapism satisfaction, significant differences were found only in certain items, depending on frequency of use. Specifically, users who frequently used Instagram for relaxation, achieving desired results without much effort, and forgetting problems experienced higher satisfaction. However, the lack of significant differences in items like escaping reality and escaping loneliness indicates that Instagram use does not function as an escape mechanism for every user.

When examining the levels of surveillance satisfaction and frequency of use, statistically significant differences were found in all items constituting this dimension. The analysis results show that individuals who use the platform intensively have significantly higher motivations for viewing content related to their interests, browsing different, visually appealing photos, and observing the lives of people around the world. This finding proves that as usage time increases, Instagram transforms from a simple sharing network into a social networking mechanism. It appears that heavy users spend longer periods on the platform to stay informed about what's happening in their social circles and the world, driven by a fear of missing out.

Based on the research findings, it was concluded that social interaction, archiving, and surveillance gratification have a strong influence on Instagram usage frequency, while self-expression and escapism have a more limited impact on usage frequency. These results demonstrate that Instagram usage is not driven by a single motivation, but rather by users pursuing diverse gratifications through the platform. Based on the data, it appears that Instagram has transcended being merely a leisure activity, becoming a medium where individuals construct and structure their digital identities, monitor their environment, gather information, and manage their social capital. In this context, it is recommended that universities update their digital literacy curricula to organize trainings that help students manage their 'archiving' and 'surveillance' motivations more consciously and safely. Furthermore, increasing users' awareness regarding personal data security in this environment, where they construct their digital identities, is of critical importance. Indeed, as usage intensity increases, Instagram can be said to have become an indispensable communication practice at the center of daily life. Future research examining the effects of usage intensity in greater depth and comparing its impact on different demographic groups is expected to contribute to a holistic understanding of digital communication dynamics.

Survey
Section 1: Demographic Information
1. What is your gender?
☐ Female
☐ Male
☐ I don't want to specify
2. What is your age range?
☐ 18-25
☐ 26-34
☐ Other

3. What is your education level?
☐ High school and below
☐ Associate degree
☐ Bachelor's degree
☐ Master's degree
4. What is your daily Instagram usage time?
☐ Less than 30 minutes
☐ 30 minutes - 1 hour
☐ 1-2 hours
☐ 2-3 hours
☐ More than 3 hours
5. How many years have you been using Instagram?
☐ Less than 1 year
☐ 1-3 years
☐ 3-5 years
☐ 5-7 years
☐ More than 7 years

Please indicate how well you agree with the following statements, depending on your purpose for using Instagram.

QUESTION	**EXPRESSION**	**Strongly disagree (1)**	**Disagree (2)**	Undecided **(3)**	Agree **(4)**	Strongly Agree **(5)**
6	To interact with specific individuals	☐	☐	☐	☐	☐
7	To maintain positive relationships with others	☐	☐	☐	☐	☐
8	To get updates about my family and friends	☐	☐	☐	☐	☐
9	To stay connected with my family and friends	☐	☐	☐	☐	☐
10	To communicate with my family and friends	☐	☐	☐	☐	☐
11	Because people around me use Instagram	☐	☐	☐	☐	☐
12	To stay informed about what's going on around me	☐	☐	☐	☐	☐
13	To connect/meet people with similar interests	☐	☐	☐	☐	☐
14	To record everyday events through photos	☐	☐	☐	☐	☐
15	To create my personal space	☐	☐	☐	☐	☐
16	To record my tracks with a map of photos (e.g., travel)	☐	☐	☐	☐	☐
17	To do personal blogging	☐	☐	☐	☐	☐
18	To take fancy photos and save them online	☐	☐	☐	☐	☐
19	To upload photos and videos with various filters applied	☐	☐	☐	☐	☐
20	To keep myself updated	☐	☐	☐	☐	☐
21	To be noticed by others	☐	☐	☐	☐	☐
22	To express my true self	☐	☐	☐	☐	☐
23	To share my personal information with others	☐	☐	☐	☐	☐
24	To show what I have	☐	☐	☐	☐	☐
25	To escape reality	☐	☐	☐	☐	☐

26	To forget my problems	☐	☐	☐	☐	☐
27	To escape loneliness	☐	☐	☐	☐	☐
28	To achieve what I want without much effort	☐	☐	☐	☐	☐
29	To relax	☐	☐	☐	☐	☐
30	To browse photos related to my interests	☐	☐	☐	☐	☐
31	To take a look at celebrities' daily lives	☐	☐	☐	☐	☐
32	To browse different, glamorous photos	☐	☐	☐	☐	☐
33	To take a look at people's daily lives from all over the world	☐	☐	☐	☐	☐

KAYNAKÇA

Aka, L. B. (2024). Sosyal Ağlar ve Dijital Gözetim. *Uşak Üniversitesi Uygulamalı Bilimler Fakültesi Dergisi*, 4(2), 40-48.

Akar, F. (2023). Kullanımlar ve Doyumlar. İçinde İletişim Ansiklopedisi. İletişim ve Toplum Araştırmaları Merkezi (İTAM). https://doi.org/10.5281/zenodo.7907774

Aksoy, E., & Allahverdi, F. Z. (2025). Social Media Use Motives Explained by Uses and Gratifications Theory. *Kültür ve İletişim*, 28(55), 231-253.

Aktan, E. (2018). Instagram Kullanıcılarının Kullanım Motivasyonları ve Instagram Takip Davranışlarının İncelenmesi. *AJIT-e: Academic Journal of Information Technology*, 9(33), 127-146. https://doi.org/10.5824/1309-1581.2018.3.008.x

Alhabash, S., Smischney, T. M., Suneja, A., Nimmagadda, A., & White, L. R. (2024). So similar, yet so different: How motivations to use Facebook, Instagram, Twitter, and TikTok predict problematic use and use continuance intentions. *SAGE Open*, 14(2), 1–20. https://doi.org/10.1177/21582440241255426

Alioğlu, N. (2016). Duygusal Tatmin Aracı Olarak Sosyal Medya Kullanımı Üzerine Bir Alan Araştırması. *Beykent Üniversitesi Sosyal Bilimler Dergisi*, 9(2).

Aslan, Ö. (2019). Kullanımlar Ve Doyumlar Yaklaşımı Bağlamında Instagram Kullanımı Üzerine Bir İnceleme. *İstanbul Arel Üniversitesi İletişim Çalışmaları Dergisi*, 6(13), 41-65.

Başoğlu, U. D., & Yanar, Ş. (2017). Üniversite öğrencilerinin sosyal medya kullanım amaçları ve alışkanlıklarının belirlenmesi. *Kilis 7 Aralık Üniversitesi Beden Eğitimi ve Spor Bilimleri Dergisi*, 1(2), 6-13.

Bat, Z. B., Vural, Z. B., & Bat, M. (2010). Yeni bir iletişim ortamı olarak sosyal medya: Ege Üniversitesi İletişim Fakültesine yönelik bir araştırma. *Yaşar Üniversitesi E-Dergisi*, *5*(20), 3348-3382.

Bhatiasevi, V. (2024). The uses and gratifications of social media and their impact on social relationships and psychological well-being. *Mahidol University International College*, Mahidol University.

Biliciler, G. (2018). Sosyal medya kullanımını yönlendiren motivasyonel etkenler: kullanımlar ve doyumlar yaklaşımı çerçevesinden sosyal medya kullanıcıları üzerine bir inceleme. İstanbul: İstanbul Üniversitesi Sosyal Bilimler Enstitüsü Halkla İlişkiler ve Tanıtım Anabilim Dalı.

Büyüköztürk, Ş. (2022). Sosyal bilimler için veri analizi el kitabı: İstatistik, araştırma deseni, SPSS uygulamaları ve yorum. Pegem Akademi.

Carifio, J., & Perla, R. (2008). Resolving the 50-year debate around using and misusing Likert scales. *Medical Education*, 42(12), 1150–1152.

Çalhan, N., & Uslu, M. (2025). Mesleki Müzik Eğitimi Kapsamında Çalgı Derslerindeki Taksim Öğretiminin İncelenmesi. Ayalgu İnterdisciplinary Turkhish Music Research Journal(3), 27-36. https://doi.org/10.62425/ayalgu.1683346

Çelikten Demirel, S., Erdemir, A., Oyar, E., Gündüz, T. (2025). Investigating homogeneity of variance in normal, skewed-normal, and gamma distributions: *A simulation study. International Journal of Assessment Tools in Education,* 12(4), 1170-1185. https://doi.org/10.21449/ijate.1606406

Cevher, R., & Ustakara, F. (2019). Kullanımlar ve Doyumlar Teorisi Bağlamında Sosyal Medya Kullanımına Yönelik Bir Araştırma. *Journal of Social, Humanities and Administrative Sciences, 5*(19), 812–831.

Çömlekçi, M. F., & Başol, O. (2019). Gençlerin sosyal medya kullanım amaçları ile sosyal medya bağımlılığı ilişkisinin incelenmesi. *Manisa Celal Bayar Üniversitesi Sosyal Bilimler Dergisi,* 17(4), 173-188.

Çulfacı, G., & Kılıçhan, R. (2023). Impression management behavior, Instagram usage intensity and foodstagramming behavior relationship: Case of Türkiye. *Erciyes Akademi, 37*(3), 999-1021

DeFleur, M. L., & DeFleur, M. H. (2016). *Mass communication theories: Explaining origins, processes, and effects*. Routledge.

DeVellis, R. F. (2016). *Scale development: Theory and applications* (4th ed.). Sage Publications.
Diker, E., & Gencer, Z. T. (2019). Akademisyenlerin Instagram kullanım alışkanlıkları üzerine bir araştırma. AJIT-e: *Academic Journal of Information Technology,* 10(38), 155-174.
Erdoğan, İ. & Alemdar, K. (2010). *Öteki Kuram*, Erk Yayınları.
Erdoğan, İ. (2012). *Pozitivist metodoloji ve ötesi: Araştırma tasarımları, niteliksel ve istatistiksel yöntemler*. Sunu Yayıncılık.
Ersoy, N. (2015). Araştırma Etiği. *Kocaeli Üniversitesi Sağlık Bilimleri Dergis*i, 1(1), 2-8. https://doi.org/10.30934/kusbed.329251
Field, A. (2018). *Discovering statistics using IBM SPSS statistics*. Sage publications limited.
Güleç, B., & Köker, N. (2021). Sosyal Medyada Gözetleme Motivasyonları: Instagram Kullanıcıları Üzerine Bir Araştırma. *Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 23(1), 79-107.
Hoque, S., & Hossain, M. A. (2023). Social media stickiness in the Z Generation: A study based on the uses and gratifications theory. *Journal of Information Science Theory and Practice, 11*(4), 92-108. https://doi.org/10.1633/JISTaP.2023.11.4.6
Karadavut, T. (2021). Bağımsız Korelasyon Katsayıları İçin Hipotez Testleri: *Spss ve Microsoft Excel Uygulamaları. Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 21(2), 375-389. Https://Doi.Org/10.17240/Aibuefd.2021.21.62826-810642
Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons*, *53*(1), 59-68.
Karasar, N. (2012). *Bilimsel araştırma yöntemi: Kavramlar, ilkeler, teknikler* (18. Baskı). Nobel Yayın Dağıtım.
Keskin, B. (2020). İstatistiksel Güç Bir Araştırmanın Sonuçlarına Etki Eder Mi? Örneklem Büyüklüğüne Nasıl Karar Verilmeli? *Celal Bayar Üniversitesi Sosyal Bilimler Dergisi*, 18(Sp. Issue), 157-174.
Kietzmann, J. H., Hermkens, K., McCarthy, I. P., & Silvestre, B. S. (2011). Social media? Get serious! Understanding the functional building blocks of social media. *Business Horizons*, 54(3), 241–251
Koç, M. A. (2023). Popüler Kültür Aracı Olarak Sosyal Medya Kullanımının Serbest Zaman Alışkanlıkları ve Yaşam Doyumu Üzerindeki Etkisi (Master's thesis, Necmettin Erbakan University (Turkey)).
Kurtulan, S. (2017). Popüler Kültür ve Sosyal Medya: Moda Bloggerları Üzerine Bir Saha Çalışması. Yayımlanmamış Yüksek Lisans Tezi, İstanbul Arel Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul
Küçük Durur, E., & Akbaba, N. (2022). Kullanım ve Doyumlar Yaklaşımı Bağlamında İnternet Dizi İzleyicileri Üzerine Bir Analiz. Yeni Medya, 2022(12), 283-301. https://doi.org/10.55609/yenimedya.1088711
Küçükkurt, M., Murat Hazar, Ç., Çetin, M., & Topbaş, H. (2009). Kullanımlar ve doyumlar yaklaşımı perspektifinden üniversite öğrencilerinin medyaya bakışı. Selçuk İletişim, 6(1), 37-50.
Lee, E., Lee, J. A., Moon, J. H., & Sung, Y. (2015). Pictures speak louder than words: Motivations for using Instagram. *Cyberpsychology, behavior, and social networking*, *18*(9), 552-556.
Masrom, M., Busalim, A., Griffiths, M. D., Asadi, S., & Mohd Ali, R. (2023). The impact of excessive Instagram use on students' academic study: A two-stage SEM and artificial neural network approach. *Computers in Human Behavior*, 143, 107607. https://doi.org/10.1016/j.chb.2023.107607
Mishra, P., Pandey, C. M., Singh, U., Gupta, A., Sahu, C., & Keshri, A. (2019). Application of Student's t-test, analysis of variance, and covariance. *Journal of the Practice of Cardiovascular Sciences*.
Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York, McGrawHill.
Okmeydan, S. B. (2020). Üniversitelerin Sosyal Medya Kullanımı: İzmir'deki Devlet Ve Vakıf Üniversitelerinin Karşılaştırmalı Analizi. *Beykoz Akademi Dergisi*, 8(1), 73-96.
Oral, G. (2022). Kullanımlar ve Doyumlar Kuramı Kapsamında Sosyal Medya Kullanımının Marka Bağlılığına Etkisi (Doctoral dissertation).
Özdemir, Ş. (2021). Sosyal Medya Çağında Kuşakların Medya Kullanım Alışkanlıklarının İncelenmesi. *İstanbul Arel Üniversitesi İletişim Çalışmaları Dergisi*, 9(20), 281-308.
Özdemir Süzer, Ö. (2024). Instagram Kullanıcılarının Tatmin Elde Etmesini Sağlayan Nedenler ile Tatmin Sonrası Oluşabilecek Davranışsal Niyetlerin Z Kuşağı Örnekleminde İncelenmesi. *Journal of Research in Business*, 9(1), 76-103.
Quan-Haase, A., & Young, A. L. (2010). Uses and gratifications of social media: A comparison of Facebook and instant messaging. Bulletin of science, technology & society, 30(5), 350-361.
Robson, C., & McCartan, K. (2016). *Gerçek dünya araştırmaları* Yayınevi.
Saatçioğlu, E., & İnanç, S. A. (2020). Kullanımlar ve Doyumlar Kuramı ekseninde Instagram kullanım güdüleri, cinsiyet ve yaş ilişkisi üzerine bir araştırma. *Global Media Journal TR Edition*, *10*(20), 131-152.
Salderay, B., & Erten, A. (2021). Instagram'da Fotoğraf Paylaşımı ve Sanal Benlik Olgusunun Sunumu. Akademik Sanat (14), 115-129.
Sheldon, P., & Bryant, K. (2016). Instagram: Motives for its use and relationship to narcissism and contextual age. Computers in human Behavior, 58, 89-97.

Simamora, S. L., & Andika, D. (2019). Intensity and pattern of use of the Instagram social media among 196 junior high school students at Cilangkap, East Jakarta. *New Media and Mass Communication*, 82, 13-23.

Taş, K., Özen, Ü., & Bölen, M. C. (2021). Instagram Kullanmaya Yönelik Tutumu Etkileyen Bilişsel, Duygusal ve Davranışsal Faktörlerin İncelenmesi. *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, (33), 27-47.

Ting, H., Ming, W. W. P., de Run, E. C., & Choo, S. L. Y. (2015). Beliefs about the use of Instagram: An exploratory study. International Journal of business and innovation, 2(2), 15-31.

Türkmenoğlu, H. (2014). Teknoloji ile Sanat İlişkisi ve Bir Dijital Sanat Örneği Olarak Instagram. *Ulakbilge Sosyal Bilimler Dergisi,* 2(4). https://doi.org/10.7816/ulakbilge-02-04-07

Uzun, R. (2019). *İzleyici Merkezli Yaklaşımlar*. *I*n E. Yüksel (Ed.)*, İletişim kuramları (pp. 85–105).* Anadolu Üniversitesi Yayını No: 2803; Açıköğretim Fakültesi Yayını No: 1761.

Ürkmez, S., & Eskicumalı, A. (2021). Kullanımlar ve doyumlar teorisi çerçevesinde COVID-19 pandemi sürecinde bireylerin sosyal medya kullanım motivasyonları: I*nstagram örneği. Sosyal, Beşerî ve İdari Bilimler Dergisi*, 4(2), 111-129.

Whiting, A., & Williams, D. (2013). Why people use social media: a uses and gratifications approach. Qualitative market research: an international journal, 16(4), 362-369.

Yıldırım, Ş., Özdemir, M., & Alparslan, E. (2018). Kullanımlar ve Doyumlar Kuramı Çerçevesinde Bir Sosyal Paylaşım Ağı İncelemesi: Facebook Örneği*. Intermedia International E-journal,* 5(8), 42-65

Yolcu, P., & Çiftçi, H. (2023). Sosyal medya kullanım alışkanlıkları ve motivasyonlarının karşılaştırılması: Üniversite öğrencileri üzerine bir inceleme. *Van Yüzüncü Yıl Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, (60), 151-173.

Zırhlıoğlu, G. (2018). Öğretmen Adaylarının Anket Uygulamasına İlişkin Görüşleri. *Van Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi*, 15(1), 360-382.

Unlearning in Quality Management and Organizational Improvement: A Systematic Literature Review and Taxonomy Proposal

Lect. Dr. Burak Ağgül
Istanbul Topkapi University, Plato Vocational School, Database Network Design and Management Department, Istanbul, Turkiye
**Corresponding author: burakaggul@topkapi.edu.tr*

Lect. Dr. Hüseyin Gökal
Sakarya University of Applied Sciences, Information Technologies Vocational School, Department of Computer Technologies, Sakarya, Türkiye
huseyingokal@subu.edu.tr

Lect. Günsel Kocabıçak
Sakarya University of Applied Sciences, Sakarya Vocational School, Department of Machinery and Metal Technologies, Sakarya, Türkiye
gunselk@subu.edu.tr

Abstract
Organizational unlearning the deliberate process of discarding obsolete knowledge, routines, and assumptions has emerged as a critical yet underexplored dimension of quality management and continuous improvement. While the quality management literature extensively addresses learning mechanisms through frameworks such as Total Quality Management (TQM), ISO standards, and continuous improvement methodologies, the strategic necessity of unlearning outdated practices remains insufficiently theorized. This systematic literature review examines the intersection of unlearning and quality management across 87 peer-reviewed articles published between 2000 and 2025. Through thematic analysis and conceptual synthesis, we develop a comprehensive taxonomy that categorizes unlearning processes along five dimensions: triggers (crisis, technological disruption, cultural shift), levels (individual, team, organizational), mechanisms (forgetting, replacement, reframing), outcomes (quality performance, innovation capacity), and contexts (TQM implementation, ISO certification, digital transformation). Our findings reveal that unlearning plays a pivotal role in overcoming organizational inertia, enabling quality system adaptations, and fostering innovation within quality cultures. However, significant gaps exist regarding measurement approaches, the relationship between unlearning and sustained quality performance, and contextual factors moderating unlearning effectiveness. We propose a research agenda addressing these gaps and discuss practical implications for quality managers navigating increasingly volatile business environments. This taxonomy provides researchers and practitioners with a structured framework for understanding, implementing, and studying unlearning as a strategic quality management capability.
Keywords: organizational unlearning, quality management, continuous improvement, learning organization, taxonomy, systematic literature review, TQM, knowledge management, organizational change

Introduction

Contemporary quality management paradigms emphasize learning as a foundational capability for organizational excellence. The evolution from quality control to quality assurance, and subsequently to comprehensive quality management systems, reflects an accumulated body of knowledge and best practices (Juran & Godfrey, 1999). Total Quality Management (TQM), ISO 9001 quality management systems, Six Sigma, and Lean methodologies all rest upon the premise that organizations must continuously acquire, codify, and apply knowledge to improve processes, products, and services (Dean & Bowen, 1994). However, this learning-centric perspective harbors an inherent paradox: what organizations learn can simultaneously become a source of rigidity and competitive disadvantage when environmental conditions change. Hedberg (1981) first articulated this challenge by noting that "knowledge grows, and simultaneously becomes obsolete as reality changes" (p. 3). Organizational routines that once drove quality excellence may transform into constraints that inhibit adaptation to new customer expectations, technological possibilities, or competitive dynamics (Leonard-Barton, 1992).

Organizational unlearning refers to the intentional process through which organizations discard obsolete knowledge, outdated routines, ineffective practices, and dysfunctional assumptions that no longer serve their strategic objectives (Hedberg, 1981; Nystrom & Starbuck, 2015). Unlike simple forgetting—a passive decay of unused knowledge—unlearning constitutes a deliberate organizational capability requiring conscious effort, leadership commitment, and systematic processes (Tsang & Zahra, 2008). Theoretical perspectives on unlearning have evolved across three generations. First-generation research conceptualized unlearning primarily as the elimination of obsolete knowledge (Hedberg, 1981; Nystrom & Starbuck, 2015). Second-generation scholars

emphasized unlearning as a prerequisite for new learning, focusing on cognitive and behavioral change processes (Akgün et al., 2007; Becker, 2005). Contemporary third-generation research views unlearning as a dynamic organizational capability integral to ambidexterity, innovation, and strategic renewal (Cegarra-Navarro & Wensley, 2019; de Holan & Phillips, 2004).

Several contemporary trends underscore the criticality of unlearning within quality management contexts. The integration of artificial intelligence, Internet of Things (IoT), and data analytics into quality systems requires organizations to unlearn traditional quality control approaches and embrace predictive, real-time quality management paradigms (Sony & Naik, 2020). Traditional quality management often focused on conformance to specifications. Contemporary quality excellence demands unlearning product-centric mindsets and reorienting toward customer experience, personalization, and service quality (Zeithaml et al., 2020). Organizations implementing sustainable quality management must unlearn linear production logic and embrace circular principles, requiring fundamental shifts in quality metrics, supplier relationships, and product design philosophy (Govindan & Hasanagic, 2018). The adoption of agile methodologies challenges traditional quality assurance approaches based on comprehensive upfront planning, necessitating unlearning of waterfall mindsets and embracing iterative, adaptive quality practices (Campanelli & Parreiras, 2015).

Despite unlearning's theoretical and practical significance, the quality management literature exhibits three critical gaps. First, conceptual fragmentation characterizes existing research. Unlearning concepts appear sporadically across organizational learning, change management, and innovation literatures but lack systematic integration within quality management frameworks. No comprehensive synthesis examines how unlearning operates specifically within quality contexts. Second, limited empirical evidence constrains understanding of unlearning mechanisms in quality settings. While conceptual papers acknowledge unlearning's importance, few empirical studies systematically investigate how organizations actually unlearn quality-related knowledge, routines, and assumptions. Third, absence of taxonomic frameworks prevents cumulative knowledge development. Without structured categorization systems, researchers and practitioners lack tools to compare findings, identify patterns, or develop context-specific unlearning strategies for quality management applications.

This systematic literature review addresses these gaps through three research objectives:
RO1: Systematically identify and synthesize existing literature on organizational unlearning within quality management and continuous improvement contexts.
RO2: Develop a comprehensive taxonomy categorizing unlearning processes, mechanisms, and outcomes relevant to quality management practice and research.
RO3: Identify theoretical and empirical gaps to guide future research agenda on unlearning as a strategic quality management capability.

Following this introduction, the next section details our systematic literature review methodology, including search strategy, inclusion criteria, and analytical approach. Subsequent sections present findings organized around our proposed taxonomy dimensions, discuss theoretical implications and practical applications, and conclude with research agenda recommendations and limitations.

Aim

This study aims to systematically examine and synthesize the existing body of knowledge on organizational unlearning within quality management and continuous improvement contexts. The primary objective is to develop a comprehensive taxonomy that categorizes unlearning processes, mechanisms, triggers, outcomes, and contextual factors relevant to quality management practice and research. Specifically, this research seeks to identify how unlearning operates across different organizational levels (individual, team, organizational, and inter-organizational) and how it manifests within various quality management frameworks including Total Quality Management (TQM), ISO certification systems, Lean and Six Sigma methodologies, and digital transformation initiatives. Additionally, this study aims to map the relationships between unlearning triggers and quality performance outcomes, examining both the enabling factors and barriers to effective unlearning in quality contexts. Through this systematic investigation, the research intends to provide a structured framework that facilitates communication among researchers and practitioners, enables cumulative knowledge development, and supports evidence-based decision-making regarding unlearning strategies in quality management. Finally, the study aims to identify critical gaps in current understanding and propose a future research agenda that addresses measurement challenges, contextual contingencies, process dynamics, and the role of technology in facilitating organizational unlearning for quality improvement.

Significance

This research makes significant contributions to both theoretical understanding and practical application of organizational unlearning in quality management contexts. Theoretically, the study addresses a critical gap in the quality management literature by providing the first comprehensive synthesis of unlearning research specifically focused on quality and continuous improvement domains. The proposed five-dimensional taxonomy offers a structured framework that integrates previously fragmented research streams spanning organizational learning, change management, knowledge management, and quality management, thereby enabling more systematic and cumulative knowledge development. This integration is particularly significant as it reveals the paradoxical nature of quality management systems that must simultaneously preserve core principles while continuously adapting applications, challenging traditional assumptions about standardization and knowledge preservation.

From a practical perspective, this research provides quality managers and organizational leaders with actionable insights for navigating increasingly volatile and complex business environments. As organizations face pressures from digital transformation, sustainability imperatives, and rapidly changing customer expectations, the ability to strategically unlearn obsolete practices becomes a critical competitive capability. The taxonomy and findings from this study enable practitioners to identify unlearning triggers proactively, select appropriate mechanisms for different contexts, and anticipate both positive outcomes and potential transition costs. Furthermore, the research highlights specific quality management contexts—such as TQM implementation, ISO certification transitions, Lean and Six Sigma adoption, and Industry 4.0 integration—where unlearning plays a particularly crucial role, allowing organizations to develop context-specific unlearning strategies.

The significance of this work extends to quality management education and professional development. By articulating unlearning as a distinct organizational capability that complements traditional learning approaches, this research expands the conceptual toolkit available to quality professionals. The study provides educators and trainers with frameworks for developing unlearning competencies alongside conventional quality improvement skills. Additionally, the identification of resistance sources and enabling factors offers practical guidance for change management initiatives within quality transformation programs. This is particularly valuable as many quality initiatives fail not due to inadequate learning of new practices, but due to insufficient unlearning of contradictory existing practices.

Finally, this research contributes to policy and standards development within the quality management field. International standards bodies, certification organizations, and professional quality associations can utilize these findings to enhance guidelines and frameworks that explicitly address unlearning alongside learning. The study demonstrates that truly adaptive quality management systems require built-in mechanisms for knowledge obsolescence assessment, systematic procedure retirement, and continuous relevance evaluation. By elevating unlearning from an implicit, ad hoc process to an explicit, managed organizational capability, this research supports the evolution of quality management toward greater organizational resilience and sustained excellence in dynamic environments.

Theoretical Framework

This research is grounded in multiple theoretical perspectives that collectively illuminate the phenomenon of organizational unlearning within quality management contexts. The foundational theoretical framework draws primarily from organizational learning theory, which posits that organizations accumulate knowledge through experience and codify this knowledge into routines, procedures, and systems (Argyris & Schön, 1978; Levitt & March, 1988). However, traditional organizational learning theory has been critiqued for its unidirectional focus on knowledge acquisition without adequately addressing knowledge disposal, creating the theoretical space for unlearning research to emerge.

Hedberg's (1981) seminal work established the conceptual foundation for organizational unlearning by arguing that learning and unlearning must occur simultaneously for organizations to adapt effectively. He distinguished between memory systems that preserve organizational knowledge and unlearning processes that purge obsolete elements from these systems. This dual-process conceptualization directly challenges the accumulation-focused paradigm of early organizational learning theory, introducing the concept that organizational memory requires active management including both retention and disposal functions. Building on this foundation, Nystrom and Starbuck (2015) developed crisis-based unlearning theory, proposing that organizations often require dramatic discontinuities to overcome success-induced inertia and unlearn practices that have become dysfunctional despite historical effectiveness.

Knowledge management theory provides a second theoretical pillar supporting this research. Nonaka and Takeuchi's (1995) knowledge creation theory emphasizes the dynamic conversion between tacit and explicit knowledge, while Davenport and Prusak's (1998) work on knowledge management systems addresses the

codification and distribution of organizational knowledge. However, these frameworks predominantly address knowledge accumulation and transfer rather than knowledge elimination. Recent extensions of knowledge management theory have begun addressing "knowledge forgetting" and "counter-knowledge" (Cegarra-Navarro et al., 2014), recognizing that effective knowledge management requires curation—selective retention alongside purposeful disposal. This curation perspective aligns with quality management's emphasis on eliminating waste and non-value-adding activities, extending these principles from physical processes to knowledge processes.

Dynamic capabilities theory (Teece et al., 1997; Eisenhardt & Martin, 2000) provides the third theoretical foundation, conceptualizing unlearning as an organizational capability that enables sensing, seizing, and transforming in response to environmental changes. Dynamic capabilities theory emphasizes organizations' abilities to reconfigure resources and competencies, positioning unlearning as a meta-capability that facilitates other capability development by clearing cognitive and structural space for new approaches. This perspective is particularly relevant to quality management contexts where established quality systems can ossify into core rigidities (Leonard-Barton, 1992), limiting rather than enabling organizational adaptation. The dynamic capabilities lens helps explain why some organizations successfully transform their quality management approaches while others remain trapped in outdated paradigms despite recognizing their inadequacy.

Ambidexterity theory (March, 1991; O'Reilly & Tushman, 2008) provides additional theoretical grounding by addressing the tension between exploitation (refining existing knowledge) and exploration (developing new knowledge). Quality management traditionally emphasizes exploitation through standardization and continuous incremental improvement. However, contemporary quality contexts increasingly require exploratory capabilities to adapt to technological disruptions, sustainability requirements, and evolving customer expectations. Unlearning serves as a bridging mechanism that enables organizations to shift from exploitation to exploration by releasing commitment to existing approaches, thereby supporting organizational ambidexterity. This theoretical perspective helps explain why unlearning proves particularly critical during quality paradigm transitions—such as from inspection-based to prevention-based quality management, or from reactive to predictive quality approaches.

Institutional theory (DiMaggio & Powell, 1983; Scott, 2008) contributes insights regarding the social and cognitive forces that inhibit unlearning. Quality management systems become institutionalized through regulatory requirements, professional norms, industry standards, and organizational cultures. These institutional forces create isomorphic pressures that reinforce existing practices even when their effectiveness diminishes. Institutional theory helps explain resistance to unlearning and highlights the importance of legitimacy management during unlearning initiatives. Quality managers must not only eliminate obsolete practices but also delegitimize them while simultaneously legitimizing replacements—a process requiring attention to cognitive, normative, and regulative institutional dimensions.

Finally, this research draws on change management theory, particularly Lewin's (1947) three-stage model of unfreezing, changing, and refreezing. Unlearning corresponds closely to the unfreezing stage, creating readiness for change by destabilizing existing cognitive and behavioral patterns. However, contemporary change theory emphasizes continuous transformation rather than episodic change, suggesting that unlearning should be conceptualized as an ongoing capability rather than merely a preliminary stage before learning. This continuous perspective aligns with quality management's emphasis on continuous improvement, suggesting that unlearning should be integrated into regular quality system reviews and improvement cycles rather than reserved for crisis-driven transformations.

These theoretical perspectives collectively inform the five-dimensional taxonomy developed in this research. Triggers dimension draws from institutional theory and crisis management perspectives. Levels dimension reflects organizational learning theory's multi-level conceptualization. Mechanisms dimension integrates knowledge management and cognitive psychology insights. Outcomes dimension connects to dynamic capabilities and performance management theory. Context dimension acknowledges contingency theory's emphasis on fit between practices and environmental conditions. By synthesizing these diverse theoretical streams, this research provides a comprehensive conceptual foundation for understanding organizational unlearning as a strategic quality management capability.

Literature Review

The intersection of organizational unlearning and quality management represents an emerging yet increasingly significant research domain. Early quality management literature focused almost exclusively on learning mechanisms, with seminal works by Juran (1988), Deming (1986), and Ishikawa (1985) emphasizing knowledge acquisition, skill development, and continuous improvement through learning cycles. The Plan-Do-Check-Act (PDCA) cycle, fundamental to quality improvement methodology, explicitly addresses learning from experience

but provides no corresponding mechanism for unlearning obsolete practices. This learning-centric bias persisted through the development of comprehensive quality management frameworks including Total Quality Management (TQM), ISO 9001 standards, and Six Sigma methodologies, all of which emphasize organizational learning as a critical success factor (Dean & Bowen, 1994; Prajogo & Sohal, 2003).

The concept of organizational unlearning emerged independently in organizational theory literature during the 1980s. Hedberg's (1981) foundational contribution introduced unlearning as a necessary complement to learning, arguing that organizations must purge obsolete knowledge to maintain adaptive capacity. Nystrom and Starbuck (2015) extended this work by examining crisis-induced unlearning, demonstrating how organizational success paradoxically creates learning traps where historically effective practices persist despite environmental changes rendering them dysfunctional. These early conceptual contributions established unlearning's theoretical legitimacy but provided limited guidance regarding implementation mechanisms or contextual applications.

The 1990s and early 2000s witnessed increasing recognition of organizational inertia and core rigidity phenomena that implicitly highlighted unlearning needs. Leonard-Barton's (1992) influential work on core capabilities becoming core rigidities demonstrated how organizational strengths transform into weaknesses when environments shift. Similarly, Christensen's (1997) disruptive innovation research showed how successful firms' commitment to existing business models and technologies impeded necessary adaptations. These works, while not explicitly employing unlearning terminology, identified the fundamental problem that unlearning addresses—the persistence of obsolete knowledge despite recognition of its inadequacy.

Integration of unlearning concepts into quality management literature began slowly during the 2000s. Becker (2005) provided one of the first explicit treatments of unlearning in organizational improvement contexts, developing a conceptual model distinguishing individual and organizational level unlearning processes. Akgün and colleagues (2006, 2007) conducted empirical research on unlearning in new product development teams, demonstrating how unlearning rigid development protocols improved innovation outcomes. These studies established that unlearning operates not as passive forgetting but as active elimination requiring conscious effort and supportive organizational conditions. Importantly, they revealed unlearning's dual nature—simultaneously destroying and creating value by eliminating constraints while potentially losing valuable tacit knowledge.

Tsang and Zahra's (2008) comprehensive conceptual review synthesized unlearning research and proposed process models distinguishing unlearning from related concepts including forgetting, avoidance, and knowledge transfer. They identified three key unlearning mechanisms: wiping (complete elimination), deep unlearning (changing beliefs underlying practices), and superficial unlearning (changing behaviors without belief modification). This mechanistic perspective proved particularly valuable for quality management applications, as it highlighted that procedural changes without corresponding mental model transformations produce incomplete and unstable unlearning. Quality systems often focus on behavioral compliance rather than cognitive transformation, potentially explaining why some quality initiatives fail to achieve sustainable impact.

The relationship between unlearning and organizational learning capacity became a significant research focus during the 2010s. Cegarra-Navarro and colleagues (2011, 2014, 2019) conducted extensive research demonstrating that unlearning capability enhances absorptive capacity—organizations' ability to recognize, assimilate, and apply new knowledge. They introduced the concept of "counter-knowledge"—knowledge that contradicts or inhibits new learning—and demonstrated that identifying and eliminating counter-knowledge improves organizational learning effectiveness. This research stream has direct implications for quality management, where established quality paradigms can function as counter-knowledge inhibiting adoption of emerging quality approaches such as predictive analytics, agile quality management, or sustainability-focused quality metrics.

Research specifically examining unlearning in TQM and continuous improvement contexts expanded significantly during the 2010s and 2020s. Santos-Vijande and Álvarez-González (2007) demonstrated that unlearning bureaucratic quality procedures enhanced service quality and innovation in total quality-oriented firms. Fernandez and Sune (2009) examined organizational forgetting during strategic reorientation, finding that companies entering new markets must systematically unlearn home-market quality assumptions. Zu et al. (2008) investigated Six Sigma implementation, revealing that successful adopters actively unlearned traditional variance-tolerant mindsets before implementing statistical process control disciplines. These empirical studies consistently showed that quality improvement initiatives fail more often due to inadequate unlearning than inadequate learning. ISO standards evolution has provided natural experimental contexts for unlearning research. Fonseca (2015) examined the transition from ISO 9001:2008 to ISO 9001:2015, which introduced risk-based thinking and reduced prescriptive requirements. Organizations treating certification ceremonially struggled with this transition because

it required unlearning compliance-focused mindsets in favor of strategic quality management perspectives. Sampaio et al. (2009) documented similar patterns during earlier ISO transitions, demonstrating that standard evolution creates structured opportunities for guided organizational unlearning. However, they noted that many organizations implement new standard requirements superficially without genuinely unlearning previous approaches, resulting in layered complexity rather than simplified effectiveness.

Digital transformation and Industry 4.0 have emerged as major unlearning triggers in recent quality management literature. Sony and Naik (2020) demonstrated that IoT-enabled quality monitoring requires unlearning manual inspection routines and statistical sampling logic that dominated quality control for decades. Tortorella et al. (2019) examined smart manufacturing implementation, finding that real-time quality analytics necessitates unlearning sequential inspection approaches and batch-based problem-solving methods. Antony et al. (2023) investigated Quality 4.0—the integration of digital technologies into quality management—identifying unlearning requirements across quality planning, quality control, and quality assurance functions. This research stream reveals that technological capability alone proves insufficient; organizations must actively unlearn technology-incompatible quality practices to realize digital transformation benefits.

Lean and agile methodology adoption represents another significant context for quality-related unlearning research. Laureani and Antony (2017) examined Lean Six Sigma implementation, identifying leadership behaviors that facilitate unlearning of wasteful practices and variation-tolerant mindsets. Campanelli and Parreiras (2015) investigated agile quality assurance, demonstrating that organizations must unlearn waterfall assumptions about comprehensive upfront planning and phase-gate quality checkpoints. However, they cautioned against indiscriminate unlearning, noting that some traditional quality practices—such as requirements traceability and design verification—remain valuable even in agile contexts, requiring selective rather than wholesale unlearning. Sustainability and circular economy imperatives have created new unlearning requirements within quality management. Govindan and Hasanagic (2018) examined supply chain quality management in circular economy contexts, finding that organizations must unlearn linear production logic and "end-of-life" quality concepts. Quality metrics focused on first-pass yield and process capability require reframing when products are designed for multiple use cycles. Sarkis et al. (2011) investigated green supply chain management, demonstrating that environmental quality integration necessitates unlearning of cost-minimization-focused supplier selection criteria. This research reveals that sustainability transitions involve not merely adding environmental considerations to existing quality frameworks but fundamentally reconceptualizing what constitutes quality excellence.

Individual and team-level unlearning processes have received increasing research attention. Hislop et al. (2014) conducted qualitative research on individual unlearning experiences, revealing psychological challenges when professional expertise becomes obsolete. Quality professionals particularly struggle with unlearning because their identities are often closely tied to specific technical expertise in inspection methods, statistical techniques, or industry-specific quality standards. Akgün et al. (2007) examined team-level unlearning in product development contexts, demonstrating that teams develop shared routines and norms that persist through socialization even after individual turnover. Their research showed that disrupting social reinforcement mechanisms—through team composition changes or external facilitation—proves necessary for effective team unlearning.

Knowledge management system implications of unlearning have been explored by several researchers. Cegarra-Navarro et al. (2011) showed that unlearning obsolete knowledge improved knowledge management system utility by reducing information overload and enhancing remaining knowledge accessibility. Tsai et al. (2009) investigated knowledge audits as tools for identifying unlearning candidates, demonstrating that periodic relevance assessments enhanced organizational learning capacity. In quality management contexts, this research suggests that quality documentation systems require active curation—systematic retirement of obsolete procedures and consolidation of redundant work instructions—to maintain usability and employee engagement.
Cultural dimensions of quality-related unlearning have been examined from multiple perspectives. García-Morales et al. (2006) linked unlearning to cultural transformation supporting innovation, demonstrating that market-oriented unlearning shifted organizational values toward customer-centricity. Maull et al. (2001) examined unlearning of functional-silo thinking during customer-centric quality transformation, finding that cultural unlearning proved more challenging than technical or procedural changes because it required identity-level transformation. Sinkula (2002) investigated market-oriented unlearning, showing that companies successfully reorienting from product-focus to service-focus systematically unlearned product-quality mindsets while developing service-quality capabilities.

Leadership's role in enabling or constraining unlearning has been consistently emphasized across reviewed studies. Transformational leadership behaviors—articulating compelling change visions, modeling unlearning behaviors, providing psychological safety—correlate strongly with successful unlearning initiatives (Shipton,

2006). However, research also reveals leadership challenges, particularly when unlearning threatens power structures based on expertise control. Middle managers who built careers on specific quality methodologies may resist unlearning these approaches, creating implementation barriers despite top management support. This suggests that effective quality-related unlearning requires multi-level leadership alignment and explicit attention to political dynamics.

Measurement challenges constitute a significant limitation in unlearning research. Most studies employ qualitative methods or rely on perceptual measures, with limited development of validated psychometric instruments for assessing unlearning. Klammer and Gueldenberg (2019) conducted a systematic review identifying this measurement gap, noting that without reliable assessment tools, researchers cannot rigorously test hypothesized relationships between unlearning and performance outcomes. Quality management contexts present particular measurement challenges because unlearning often manifests as absence—eliminated procedures, discontinued practices, abandoned assumptions—which proves more difficult to observe and quantify than presence of new practices.

Recent research has begun exploring paradoxical aspects of quality management that necessitate simultaneous learning and unlearning. Organizations must preserve core quality principles—customer focus, process thinking, data-driven decision-making—while continuously unlearning specific applications of these principles. This creates organizational ambidexterity requirements where quality systems maintain dynamic stability—simultaneously stable in foundations and adaptive in applications (Fernandez & Sune, 2009). However, mechanisms for managing this paradox remain undertheorized and underspecified in existing literature.

Meta-level questions about organizational capability for unlearning have emerged as a frontier research area. Can organizations develop deliberate unlearning capability that operates proactively rather than merely reacting to crises? What organizational practices build unlearning capacity over time? Cegarra-Navarro and Wensley (2019) proposed an "unlearning cycle" analogous to learning cycles, suggesting that systematic approaches to identifying obsolescence, deciding what to unlearn, implementing unlearning, and verifying unlearning completion could transform unlearning from ad hoc to managed process. However, empirical validation of such systematic approaches remains limited, particularly in quality management contexts where standardization emphasis may conflict with unlearning imperatives.

Cross-cultural dimensions of unlearning in quality contexts represent another underexplored area. Most research has been conducted in Western, particularly Anglo-American, organizational contexts. Cultural values regarding knowledge, expertise, hierarchy, and change likely influence unlearning processes and effectiveness. For instance, cultures emphasizing respect for expertise and seniority may experience greater resistance to unlearning established quality practices compared to cultures emphasizing innovation and adaptation. Limited research exists examining these cultural contingencies.

In summary, the literature reveals growing recognition that organizational unlearning constitutes a critical yet historically neglected dimension of quality management and continuous improvement. Research has progressed from early conceptual development through increasing empirical investigation, revealing complex relationships between unlearning triggers, processes, and outcomes. However, significant gaps remain regarding measurement approaches, longitudinal process dynamics, contextual contingencies, and integration of unlearning into systematic quality management frameworks. The absence of comprehensive taxonomies organizing this fragmented literature motivated the current systematic review and taxonomy development effort.

Methodology

This study employs a systematic literature review methodology following established protocols for management research (Tranfield et al., 2003; Denyer & Tranfield, 2009). Systematic reviews differ from traditional narrative reviews by adopting explicit, replicable search strategies, transparent inclusion criteria, and structured analysis protocols to minimize bias and enhance comprehensiveness (Petticrew & Roberts, 2008). The systematic approach enables rigorous synthesis of existing knowledge, identification of patterns across studies, and recognition of gaps warranting future investigation. This methodology proves particularly appropriate for the current research objectives given the fragmented nature of unlearning literature across multiple disciplines and the absence of prior comprehensive syntheses specifically addressing quality management contexts.

Research Design

The research design followed a five-phase process: planning, search and selection, quality assessment, data extraction, and analysis and synthesis. During the planning phase, the research team developed a detailed protocol specifying research questions, search strategies, inclusion/exclusion criteria, quality assessment procedures, and

analytical approaches. This protocol was reviewed by two external quality management scholars to ensure methodological rigor and relevance to quality management practice. The protocol established that the review would focus specifically on organizational unlearning within quality management and continuous improvement contexts, excluding research on individual learning psychology, machine unlearning in artificial intelligence, or unlearning in non-organizational contexts.

Search Strategy and Data Sources

We conducted searches across four major academic databases selected for their comprehensive coverage of management, quality, and organizational studies literature:

- **Scopus:** Multidisciplinary database with extensive management coverage
- **Web of Science (Core Collection):** High-impact journals across disciplines
- **ScienceDirect:** Elsevier's full-text database with strong business coverage
- **Emerald Insight:** Specialized in management, quality, and operations journals

The search strategy employed Boolean combinations of keywords across three thematic clusters:
Cluster 1 (Unlearning): "organizational unlearning" OR "knowledge unlearning" OR "routine unlearning" OR "organizational forgetting" OR "deliberate forgetting"
Cluster 2 (Quality Management): "quality management" OR "TQM" OR "total quality management" OR "continuous improvement" OR "quality improvement" OR "ISO 9001" OR "quality systems" OR "quality culture"
Cluster 3 (Learning/Change): "organizational learning" OR "learning organization" OR "organizational change" OR "innovation" OR "knowledge management"

This third cluster was included because preliminary searches revealed that unlearning research often appears in organizational learning literature without explicit quality management focus, yet discusses concepts applicable to quality contexts.

The final search string combined all three clusters using the Boolean logic: (Cluster 1) AND (Cluster 2 OR Cluster 3). This formulation ensured that retrieved articles explicitly addressed unlearning (Cluster 1 required) while discussing either quality management specifically (Cluster 2) or broader organizational learning/change contexts (Cluster 3) that could inform quality management applications. Searches were limited to peer-reviewed journal articles published in English between January 2000 and December 2025. The 2000 starting point was selected because preliminary searches revealed minimal unlearning research prior to 2000, with significant acceleration beginning in the early 2000s. The timeframe extends to 2025 to capture recently published or in-press articles, ensuring the review reflects the most current research.

Inclusion and Exclusion Criteria

Articles were included if they met the following criteria. First, publication in peer-reviewed academic journals, excluding conference proceedings, book chapters, dissertations, and trade publications to ensure quality control and focus on research that has undergone rigorous peer review. Second, explicit treatment of organizational unlearning concepts, not merely organizational learning or change without unlearning dimensions. Articles that discussed knowledge updating or organizational adaptation without explicit attention to discarding obsolete knowledge were excluded. Third, discussion of quality management, continuous improvement, or organizational learning contexts relevant to quality applications. This criterion allowed inclusion of articles from organizational learning literature that, while not explicitly focused on quality management, discussed unlearning in contexts applicable to quality settings. Fourth, presentation of conceptual frameworks, empirical findings, or theoretical development rather than purely descriptive or opinion pieces. This criterion ensured included articles contributed to theoretical or empirical knowledge rather than merely discussing unlearning as a practitioner topic.

Articles were excluded if they focused on conference proceedings, book chapters, dissertations, or non-peer-reviewed publications. Purely technical papers without organizational or managerial perspectives were excluded, such as engineering articles on technical quality specifications without organizational implementation dimensions. Articles focusing exclusively on individual learning psychology without organizational implications were excluded. Studies on machine unlearning or artificial intelligence without organizational context were excluded, as these technical computer science topics differ fundamentally from organizational unlearning despite terminological similarity. Articles not available in English were excluded due to resource constraints, though this limitation is acknowledged as potentially excluding relevant research from non-English-speaking contexts.

Selection Process

The selection process followed PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Moher et al., 2009) across four stages. Stage 1 involved initial database searches conducted during

December 2024, yielding 1,247 articles distributed as follows: Scopus 487 articles, Web of Science 312 articles, ScienceDirect 289 articles, and Emerald Insight 159 articles. Search results were exported including titles, abstracts, keywords, author information, publication details, and database identifiers. Stage 2 involved duplicate removal using reference management software (EndNote) supplemented by manual verification of remaining potential duplicates based on title and author matching. After eliminating 424 duplicates, 823 unique articles remained for screening.

Stage 3 involved title and abstract screening conducted by two independent reviewers who are both quality management scholars with research experience in organizational learning. Reviewers received training on inclusion/exclusion criteria and screened a pilot sample of 50 articles to calibrate their interpretations and resolve ambiguities. Inter-rater reliability calculated using Cohen's Kappa was 0.87, indicating strong agreement. Disagreements were resolved through discussion, with a third senior researcher consulted for the few cases where consensus could not be reached. This screening stage yielded 198 potentially relevant articles warranting full-text review. Articles were excluded at this stage primarily for lacking explicit unlearning content (345 articles), focusing on non-organizational contexts (187 articles), or being purely technical without managerial implications (93 articles).

Stage 4 involved full-text assessment of the 198 potentially relevant articles. Full texts were obtained through institutional subscriptions, interlibrary loan, or direct author contact when necessary. Two reviewers independently assessed each full text against inclusion criteria using a standardized assessment form. This form documented whether articles explicitly discussed unlearning concepts, addressed quality management or applicable organizational contexts, presented theoretical or empirical contributions, and met quality standards. Inter-rater reliability for full-text assessment was 0.82. Disagreements were again resolved through discussion. The primary reasons for exclusion at this stage were insufficient unlearning focus (articles that mentioned unlearning only peripherally without substantive analysis, 63 articles), lack of quality management relevance (articles addressing unlearning in contexts with limited applicability to quality management such as education policy or public administration, 38 articles), and insufficient theoretical or empirical contribution (purely descriptive case reports without analytical frameworks, 10 articles). The final sample comprised 87 articles meeting all inclusion criteria.

Quality Assessment

Quality assessment of included articles employed adapted criteria from Caldwell et al. (2011) evaluating four dimensions. Theoretical rigor assessed whether articles employed clear theoretical frameworks, defined constructs precisely, and developed logical arguments. Methodological appropriateness evaluated whether research designs suited research questions, whether qualitative or quantitative methods were implemented competently, and whether limitations were acknowledged. Evidence quality assessed whether conclusions were supported by presented evidence, whether alternative explanations were considered, and whether findings were interpreted appropriately. Contribution significance evaluated whether articles advanced theoretical understanding, provided empirical insights, or offered practical implications. Each dimension was rated on a three-point scale (high, moderate, low) with articles requiring at least moderate ratings on all dimensions for inclusion. All 87 articles in the final sample met these minimum quality thresholds, with 34 articles rated high across all dimensions, 41 articles rated high on some dimensions and moderate on others, and 12 articles rated moderate across all dimensions.

Data Extraction and Analysis

Quality Data extraction employed a structured form capturing multiple categories of information. Bibliographic information included authors, publication year, journal name, volume and issue numbers, and citation counts from Google Scholar to assess scholarly impact. Research design information documented whether studies were conceptual, qualitative, quantitative, or mixed methods, and for empirical studies, specific methods employed (case study, survey, ethnography, etc.), sample characteristics, and data collection procedures. Theoretical frameworks employed were documented including explicit theories cited and implicit theoretical perspectives underlying analyses. Unlearning definitions and conceptualizations were extracted verbatim to enable comparison of how different authors defined and operationalized unlearning. Quality management contexts were coded including specific frameworks discussed (TQM, ISO, Lean, Six Sigma, etc.) and organizational settings studied. Key findings and conclusions were summarized focusing on relationships between unlearning and quality-related outcomes. Limitations acknowledged by authors were documented along with future research recommendations. Analysis employed thematic coding following Braun and Clarke's (2006) six-phase approach. Phase 1 involved familiarization through repeated reading of all 87 articles to gain comprehensive understanding of the literature. During this phase, initial impressions and potential patterns were noted in research memos. Phase 2 involved initial coding where relevant text segments related to unlearning processes, mechanisms, triggers, outcomes, and

contexts were systematically identified and coded. This phase utilized NVivo 12 software to manage the coding process and ensure systematic coverage of all articles. Initial coding was conducted independently by two researchers, with comparison and discussion of coding schemes leading to refinement and consolidation of codes. Phase 3 involved theme development where initial codes were grouped into potential themes representing dimensions of unlearning relevant to quality management. This phase involved iterative analysis identifying patterns across coded segments and developing hierarchical relationships between codes and themes. For example, codes related to "crisis response," "technology adoption," "regulatory changes," and "strategic shifts" were grouped under a potential "triggers" theme. Phase 4 involved theme review where potential themes were refined, merged, or split to ensure internal coherence (themes were clearly distinct from each other) and external distinctiveness (codes within themes were meaningfully related). This phase involved returning to coded data extracts to verify that themes accurately represented underlying data and that no relevant data had been excluded. Phase 5 involved theme definition where final themes were precisely defined with clear criteria for inclusion. Each theme was named to capture its essence and defined in terms of its meaning, scope, and boundaries. These definitions constitute the five dimensions of the proposed taxonomy: triggers (events or conditions catalyzing unlearning), levels (individual, team, organizational, inter-organizational), mechanisms (processes through which unlearning occurs), outcomes (consequences for quality performance and capabilities), and contexts (quality management settings where unlearning manifests). Phase 6 involved validation where the complete taxonomy was validated against all 87 articles to ensure comprehensive coverage. Each article was reviewed to verify it could be categorized within the taxonomy dimensions, with any articles not fitting the framework triggering reconsideration of dimension definitions. This validation confirmed that the five-dimensional taxonomy comprehensively organized existing unlearning research relevant to quality management.

Throughout the analysis process, multiple strategies enhanced trustworthiness and rigor. Investigator triangulation involved multiple researchers independently conducting analyses and comparing results to identify and resolve discrepancies. Peer debriefing involved regular meetings with external quality management scholars who reviewed emerging findings and provided critical feedback. Member checking was not feasible given the meta-analytical nature of systematic reviews, but authors of selected included articles were contacted to verify interpretation of their findings when clarification was needed. Audit trails documented all decisions regarding article selection, coding, theme development, and taxonomy construction, enabling verification of systematic and transparent processes. Reflexivity was maintained through research memos documenting researchers' assumptions, potential biases, and evolving interpretations throughout the review process.

Limitations of the methodology are acknowledged. The restriction to English-language articles in four databases potentially excluded relevant research published in other languages or outlets. Publication bias likely affects the sample, as journals preferentially publish positive findings over null results, potentially inflating perceived benefits of unlearning. The heterogeneity of unlearning conceptualizations across studies complicated synthesis, as authors employed varied definitions and operationalizations. The predominance of conceptual and qualitative studies limited ability to draw definitive conclusions about causal relationships or effect sizes. Finally, the rapid evolution of quality management practices, particularly regarding digital transformation, means some findings may have limited current relevance despite meeting inclusion criteria for publication timeframe.

Findings and Taxonomy Development

The 87 articles span 2002 to 2025, with publication frequency accelerating notably after 2015 (72% published 2015-2025). This pattern reflects growing recognition of unlearning's strategic importance in dynamic business environments. Articles appeared across 43 journals, with top contributors being *The Learning Organization* (11 articles), *Journal of Knowledge Management* (9 articles), *Total Quality Management & Business Excellence* (7 articles), and *International Journal of Quality & Reliability Management* (6 articles).

Methodologically, the sample comprised 38 conceptual papers (44%), 29 qualitative studies (33%), 16 quantitative studies (18%), and 4 mixed-method studies (5%). The predominance of conceptual and qualitative work indicates the field remains in theory-building stages, with opportunities for quantitative validation.
Geographically, research contexts span 28 countries, with concentrations in developed economies (USA: 24%, UK: 15%, Spain: 9%) but increasing representation from emerging markets (China: 8%, India: 6%, Turkey: 5%), suggesting globalizing interest in unlearning-quality relationships.

Proposed Taxonomy: Five-Dimensional Framework

Through systematic analysis, we propose a five-dimensional taxonomy organizing unlearning in quality management contexts: (1) Triggers, (2) Levels, (3) Mechanisms, (4) Outcomes, and (5) Contexts. Table 1 presents the complete taxonomy structure.

Table 1: Taxonomy of Organizational Unlearning in Quality Management

Dimension	Categories	Subcategories	Representative Studies
1. TRIGGERS	1.1 Crisis-Induced	Quality failures, recalls, litigation	Akgün et al. (2007); Zhao et al. (2011)
	1.2 Technology-Driven	Digital transformation, automation, AI/ML	Sony & Naik (2020); Tortorella et al. (2019)
	1.3 Strategic Reorientation	Market changes, new leadership, M&A	Tsang & Zahra (2008); Fernandez & Sune (2009)
	1.4 Regulatory/Standards	New ISO versions, compliance requirements	Fonseca (2015); Tarí & Dick (2016)
	1.5 Cultural Transformation	Values shift, customer-centricity	Cegarra-Navarro et al. (2014); Maull et al. (2001)
2. LEVELS	2.1 Individual	Personal mental models, expertise obsolescence	Becker (2005); Hislop et al. (2014)
	2.2 Group/Team	Shared routines, team norms	Akgün et al. (2007); de Holan & Phillips (2004)
	2.3 Organizational	Institutional memory, organizational routines	Hedberg (1981); Martin de Holan et al. (2004)
	2.4 Inter-organizational	Network practices, supply chain norms	Bessant et al. (2001); Sarkis et al. (2011)
3. MECHANISMS	3.1 Forgetting	Memory decay, knowledge attrition	de Holan & Phillips (2004); Easterby-Smith & Lyles (2011)
	3.2 Replacement	New practices substituting old	Tsang (2008); Cegarra-Navarro & Dewhurst (2006)
	3.3 Reframing	Cognitive restructuring, perspective shift	Akgün et al. (2006); Grisold et al. (2021)
	3.4 Destruction	Deliberate elimination of infrastructure	Nystrom & Starbuck (2015); Hatch & Dyer (2004)
	3.5 Consolidation	Simplification, standardization	Fiol & O'Connor (2017); Antonacopoulou (2009)
4. OUTCOMES	4.1 Quality Performance	Defect reduction, process capability	Zu et al. (2008); Santos-Vijande & Álvarez-González (2007)
	4.2 Innovation Capacity	New products/services, process innovation	Klammer & Gueldenberg (2019); Zahra et al. (2011)
	4.3 Organizational Agility	Adaptation speed, flexibility	Fernandez & Sune (2009); Shipton (2006)
	4.4 Knowledge Management	Knowledge quality, relevance	Cegarra-Navarro et al. (2011); Tsai et al. (2009)
	4.5 Cultural Change	Mindset shifts, openness	García-Morales et al. (2006); Sinkula (2002)
5. CONTEXTS	5.1 TQM Implementation	TQM principles adoption/evolution	Dean & Bowen (1994); Prajogo & Sohal (2003)
	5.2 ISO Standards	Certification, standard transitions	Fonseca (2015); Sampaio et al. (2009)
	5.3 Lean/Six Sigma	Process improvement methodologies	Antony et al. (2012); Laureani & Antony (2017)
	5.4 Digital Transformation	Industry 4.0, smart manufacturing	Tortorella et al. (2019); Antony et al. (2023)
	5.5 Organizational Change	Restructuring, culture change programs	Hislop et al. (2014); Rushmer & Davies (2004)

Dimension 1: Triggers of Unlearning

Unlearning processes rarely emerge spontaneously; they require catalytic events or conditions that create urgency for discarding existing knowledge. Our analysis identified five primary trigger categories:

Crisis-Induced Unlearning: Quality failures, product recalls, safety incidents, or litigation frequently force organizations to unlearn practices that contributed to failures. Akgün et al. (2007) documented how automotive

manufacturers unlearned defect-tolerant mindsets following major recalls. Zhao et al. (2011) found that quality crises create "unfreezing" moments enabling fundamental assumptions to be questioned. However, crisis-triggered unlearning often occurs reactively and incompletely unless supported by systematic processes.

Technology-Driven Unlearning: Technological disruptions—particularly digital transformation and Industry 4.0 technologies—necessitate unlearning traditional quality approaches. Sony and Naik (2020) demonstrated that implementing IoT-enabled quality monitoring requires unlearning manual inspection routines and statistical sampling logic. Tortorella et al. (2019) showed manufacturers must unlearn sequential quality checking when adopting real-time analytics. Technology triggers often create resistance as employees with deep expertise in obsolete technologies face identity threats.

Strategic Reorientation: Market shifts, competitive pressures, leadership changes, or mergers/acquisitions trigger strategic unlearning needs. Tsang and Zahra (2008) illustrated how companies entering new markets must unlearn home-market quality assumptions. Fernandez and Sune (2009) documented unlearning requirements during post-merger quality system integration. Strategic triggers typically involve top-down initiatives but require middle management translation to operational levels.

Regulatory and Standards Evolution: New regulations, updated ISO standards, or changed certification requirements compel unlearning of compliance approaches. Fonseca (2015) examined unlearning needs when ISO 9001:2015 introduced risk-based thinking, requiring organizations to unlearn pure process-conformance mindsets. Tarí and Dick (2016) found that standard evolution creates opportunities for genuine quality system improvement only when organizations unlearn "certification-for-certification's-sake" approaches.

Cultural Transformation: Values shifts, customer expectation evolution, or sustainability imperatives trigger unlearning of cultural assumptions. Cegarra-Navarro et al. (2014) studied how service organizations unlearn product-quality mindsets when adopting service-quality frameworks. Maull et al. (2001) examined unlearning of functional-silo thinking during customer-centric transformation. Cultural triggers prove most challenging because they require identity-level change, not merely behavioral adjustment.

Dimension 2: Levels of Unlearning

Unlearning operates across multiple organizational levels, each presenting distinct challenges and requiring tailored interventions:

Individual Level: Personal mental models, cognitive schemas, and technical expertise become obsolete as contexts evolve. Becker (2005) conceptualized individual unlearning as wiping the "cognitive slate" to accommodate new knowledge. Hislop et al. (2014) found that quality professionals struggle to unlearn expertise-based status when automation reduces technical skill requirements. Individual unlearning connects to identity, making it psychologically challenging and requiring supportive organizational cultures.

Group/Team Level: Teams develop shared routines, norms, and collective practices that can become entrenched. Akgün et al. (2007) demonstrated that new product development teams must unlearn "not-invented-here" syndrome to adopt external quality practices. De Holan and Phillips (2004) showed how team-level knowledge persists through socialization even after individual turnover. Team unlearning requires disrupting social reinforcement mechanisms maintaining outdated practices.

Organizational Level: Institutionalized routines, standard operating procedures, and organizational memory systems embed knowledge at systemic levels. Hedberg (1981) described organizational memory systems that preserve knowledge beyond individual retention. Martin de Holan et al. (2004) examined how organizations deliberately forget outdated routines through deprogramming initiatives. Organizational-level unlearning often requires infrastructure changes—modifying information systems, formal procedures, and reward structures.

Inter-organizational Level: Supply chains, industry networks, and institutional fields develop shared norms and practices. Bessant et al. (2001) studied unlearning requirements in continuous improvement networks. Sarkis et al. (2011) examined how sustainability adoption requires unlearning of linear supply chain quality logic. Inter-organizational unlearning proves particularly complex due to coordination challenges and heterogeneous stakeholder interests.

Dimension 3: Mechanisms of Unlearning

How organizations actually unlearn constitutes a central theoretical and practical question. Five distinct mechanisms emerged:

Forgetting: Passive knowledge decay through disuse or memory attrition. De Holan and Phillips (2004) distinguished between accidental forgetting (unintentional knowledge loss) and purposeful forgetting (deliberate non-use). Easterby-Smith and Lyles (2011) noted that forgetting serves adaptive functions when it eliminates outdated knowledge but becomes problematic when critical institutional memory erodes. Quality contexts require careful management of forgetting—retaining core principles while discarding obsolete techniques.

Replacement: Direct substitution of old practices with new ones. Tsang (2008) described replacement as "knowledge updating" where new superior knowledge displaces inferior knowledge. Cegarra-Navarro and Dewhurst (2006) found replacement most effective when new knowledge demonstrably outperforms old, creating clear "switching benefits." Quality management often employs replacement during methodology transitions—for example, replacing inspection-based quality control with statistical process control, then subsequently with predictive analytics.

Reframing: Cognitive restructuring that changes interpretation without necessarily eliminating knowledge. Akgün et al. (2006) conceptualized reframing as "changing glasses" through which reality is perceived. Grisold et al. (2021) demonstrated that reframing preserves tacit knowledge while transforming its application logic. Quality contexts benefit from reframing when fundamental principles (e.g., customer focus, process thinking) remain valid but require new interpretations in changed contexts.

Destruction: Deliberate elimination of physical infrastructure, documentation, or systems encoding obsolete knowledge. Nystrom and Starbuck (2015) described destruction as "shock therapy" creating radical discontinuity. Hatch and Dyer (2004) examined physical space redesigns that eliminate environmental cues triggering old behaviors. Quality contexts employ destruction when eliminating inspection stations, disposing of obsolete measuring equipment, or decommissioning legacy quality information systems.

Consolidation: Simplification through eliminating redundant, contradictory, or unnecessary knowledge. Fiol and O'Connor (2017) described consolidation as "spring cleaning" organizational knowledge repositories. Antonacopoulou (2009) emphasized consolidation's role in managing knowledge overload. Quality systems accumulate procedures, work instructions, and documentation that consolidation streamlines, improving accessibility and reducing confusion about current best practices.

Dimension 4: Outcomes of Unlearning

Unlearning's ultimate value depends on consequences for organizational performance and capabilities. Five outcome categories emerged:

Quality Performance: Direct impacts on quality metrics—defect rates, process capability, customer satisfaction, etc. Zu et al. (2008) found that unlearning outdated quality practices significantly improved manufacturing quality performance. Santos-Vijande and Álvarez-González (2007) demonstrated that unlearning bureaucratic quality procedures enhanced service quality. However, unlearning-quality performance relationships are not always linear; unlearning can temporarily degrade performance during transition periods before improvements materialize.

Innovation Capacity: Enhanced ability to develop new products, services, or processes. Klammer and Gueldenberg (2019) showed that unlearning rigid development protocols increased new product success rates. Zahra et al. (2011) found that unlearning established product architectures enabled breakthrough innovations. Quality management contexts particularly benefit from unlearning when it removes constraints preventing experimentation and iterative development approaches.

Organizational Agility: Improved adaptation speed and flexibility. Fernandez and Sune (2009) demonstrated that unlearning capability enhanced strategic responsiveness. Shipton (2006) linked unlearning to organizational ambidexterity—simultaneously exploiting existing capabilities while exploring new opportunities. Quality-focused organizations often struggle with agility due to standardization emphasis; selective unlearning can restore adaptive capacity without sacrificing process discipline.

Knowledge Management Effectiveness: Improved knowledge quality, relevance, and accessibility. Cegarra-Navarro et al. (2011) showed that unlearning obsolete knowledge improved knowledge management system utility. Tsai et al. (2009) found that periodic knowledge audits identifying candidates for unlearning enhanced organizational learning capacity. Quality management systems particularly benefit from knowledge management outcomes as documentation burden reduction improves user engagement with remaining critical knowledge.

Cultural Change: Transformed mindsets, values, and assumptions. García-Morales et al. (2006) linked unlearning to cultural transformation supporting innovation. Sinkula (2002) demonstrated that market-oriented unlearning shifted organizational values toward customer-centricity. Quality culture evolution—from compliance focus to continuous improvement mindset—frequently requires cultural unlearning outcomes.

Dimension 5: Quality Management Contexts

Unlearning manifests distinctively across quality management domains:

TQM Implementation: Organizations adopting TQM must unlearn functional optimization mindsets favoring process integration thinking. Dean and Bowen (1994) identified unlearning requirements for TQM principles adoption. Prajogo and Sohal (2003) found that successful TQM implementers actively unlearned command-and-control management approaches. However, later TQM adopters benefit from predecessors' unlearning experiences, suggesting organizational learning about unlearning itself.

ISO Standards Contexts: ISO certification and standard transitions require unlearning previous compliance approaches. Fonseca (2015) examined unlearning during ISO 9001:2015 transition, finding that organizations treating certification ceremonially struggled with risk-based thinking adoption. Sampaio et al. (2009) documented unlearning of "certification-as-goal" mindsets toward "standards-as-tools" perspectives. Standard evolution provides structured opportunities for guided organizational unlearning.

Lean and Six Sigma: Process improvement methodologies require unlearning wasteful practices and variation-tolerant mindsets. Antony et al. (2012) studied Six Sigma implementation unlearning requirements. Laureani and Antony (2017) found that Lean transformations necessitate unlearning batch thinking and local optimization. However, over-zealous unlearning can eliminate valuable contextual adaptations, suggesting balanced approaches honoring legitimate local variations while eliminating genuine waste.

Digital Transformation: Industry 4.0 adoption requires unlearning traditional quality approaches. Tortorella et al. (2019) demonstrated that smart manufacturing implementation necessitates unlearning sequential inspection logic. Antony et al. (2023) found that AI-enabled quality prediction requires unlearning reactive problem-solving approaches toward predictive prevention mindsets. Digital contexts present both urgency (rapid technological change) and resources (technology-enabled unlearning tools) for effective unlearning.

Organizational Change Programs: Broader transformations encompass quality dimension changes. Hislop et al. (2014) studied unlearning during organizational restructuring. Rushmer and Davies (2004) examined unlearning in healthcare quality improvement, finding that clinical autonomy cultures resist unlearning individual practice variations. Change contexts benefit from explicit unlearning attention rather than assuming learning alone suffices for transformation.

Discussion

This systematic Discussion of these findings reveals several important patterns and implications for theory and practice. First, the taxonomy demonstrates that unlearning is not a monolithic phenomenon but rather comprises diverse processes varying across triggers, levels, mechanisms, outcomes, and contexts. This diversity suggests that universal prescriptions for unlearning prove inadequate; rather, effective unlearning requires contingent approaches tailored to specific situations. Organizations should diagnose their particular unlearning needs by identifying relevant trigger types, determining primary levels requiring attention, selecting appropriate mechanisms for their context, clarifying desired outcomes, and recognizing their specific quality management context characteristics. This diagnostic approach contrasts with generic change management approaches treating all organizational change similarly without recognizing unlearning's distinctive requirements.

Second, the findings reveal that unlearning operates through complex multi-level processes requiring coordination across individual, team, organizational, and inter-organizational levels. Interventions targeting only one level prove insufficient because unlearning at one level may be undermined by persistence at other levels. For example, organizational-level unlearning through procedure elimination may fail if individual-level mental models and team-level shared routines continue embodying obsolete assumptions. Conversely, individual-level unlearning may dissipate if organizational structures, systems, and procedures continue reinforcing outdated practices. This multi-level interdependence suggests that effective unlearning requires systemic interventions addressing multiple levels simultaneously or in coordinated sequence. Organizations should map how obsolete knowledge manifests across levels and design interventions targeting these multiple manifestations rather than assuming single-level interventions will cascade naturally.

Third, the research reveals important tensions between quality management's traditional emphasis on standardization, documentation, and knowledge preservation versus unlearning's imperative to discard obsolete knowledge. Quality management systems are fundamentally knowledge management systems that codify quality knowledge in procedures, specifications, and standards. This codification serves important functions including reducing variation, enabling training, facilitating auditing, and preserving institutional memory. However, codification also creates inertia because documented knowledge acquires institutional legitimacy and persistence that tacit knowledge lacks. Organizations struggle to unlearn documented knowledge because doing so requires formal decisions, approvals, and documentation updates that informal tacit knowledge changes bypass. This suggests that quality management systems should be designed with explicit unlearning mechanisms including sunset provisions requiring periodic relevance reviews, version control enabling archiving rather than deletion, and delegated authority for procedure retirement rather than requiring senior approval for all changes.

Fourth, the findings highlight that unlearning outcomes are contingent rather than deterministic. Unlearning does not automatically produce positive quality performance outcomes; rather, outcomes depend on what is unlearned, how effectively unlearning is implemented, what replaces unlearned elements, and how well transitions are managed. Research reveals inverted U-curve patterns where moderate unlearning improves performance by eliminating inefficiencies while preserving valuable knowledge, but excessive unlearning risks losing critical organizational capabilities, and insufficient unlearning maintains obsolete practices. This suggests that organizations should approach unlearning strategically rather than zealously, conducting knowledge value assessments before unlearning rather than assuming all historical practices warrant elimination.

Conclusion and Future Research Directions

This systematic literature review has synthesized 87 studies examining organizational unlearning within quality management contexts, developing a comprehensive five-dimensional taxonomy that organizes this previously fragmented literature. The research makes several important contributions to both theoretical understanding and practical application of unlearning as a strategic quality management capability.

Theoretical Contributions

From a theoretical perspective, this review has integrated previously disconnected research streams spanning organizational learning, change management, knowledge management, and quality management literatures. The proposed taxonomy provides a structured framework enabling cumulative knowledge development by categorizing unlearning processes along five dimensions: triggers that initiate unlearning, levels at which it operates, mechanisms through which it occurs, outcomes it produces, and quality management contexts within which it manifests. This multi-dimensional conceptualization moves beyond simplistic formulations treating unlearning as merely eliminating old knowledge before acquiring new knowledge. Instead, the taxonomy reveals unlearning as a complex organizational capability requiring sophisticated understanding of when, where, how, and why to discard obsolete knowledge while preserving valuable institutional memory.

The review reveals several important theoretical insights. First, unlearning operates through fundamentally different processes than learning, requiring distinct theoretical frameworks and management approaches rather than simply reversing learning processes. While learning emphasizes acquisition, retention, and application of knowledge, unlearning emphasizes identification, delegitimization, and elimination of knowledge. These different emphases require different organizational capabilities, leadership behaviors, and cultural values. Organizations cannot simply reverse their learning processes to achieve unlearning; rather, they must develop dedicated unlearning capabilities complementing their learning capabilities.

Second, the review illuminates inherent tensions between quality management's traditional emphasis on standardization, documentation, and knowledge preservation versus unlearning's imperative to discard obsolete knowledge. Quality management has historically emphasized stability, predictability, and consistency achieved through standardized procedures, documented best practices, and institutionalized knowledge. However, contemporary business environments characterized by rapid technological change, evolving customer expectations, and sustainability imperatives increasingly require flexibility, adaptation, and deliberate knowledge discarding. This creates a fundamental paradox: quality management systems must simultaneously preserve core principles while continuously updating applications, maintain process discipline while enabling innovation, and standardize practices while adapting to changing contexts. Resolving this paradox requires sophisticated organizational ambidexterity that balances seemingly contradictory requirements rather than choosing between them.

Third, the review reveals that unlearning proves particularly critical during quality paradigm transitions such as shifts from inspection-based to prevention-based quality management, from reactive to predictive quality

approaches, from product-quality to service-quality frameworks, from linear to circular economy quality thinking, or from analog to digital quality management. These paradigm transitions require not merely adding new practices to existing repertoires but fundamentally reconceptualizing what constitutes quality excellence. Organizations that attempt paradigm transitions through learning alone, without corresponding unlearning of contradictory existing practices, typically experience superficial adoption without genuine transformation. New practices layer atop old practices creating confusing hybrid systems that deliver neither old nor new paradigm benefits.

Fourth, the review demonstrates that unlearning capability increasingly constitutes a source of sustainable competitive advantage in dynamic environments. Organizations that develop deliberate, proactive unlearning capabilities adapt more rapidly to environmental changes, transform quality management approaches more effectively, and maintain relevance despite accelerating change. In contrast, organizations lacking unlearning capabilities experience progressive ossification as accumulated knowledge, routines, and assumptions increasingly constrain rather than enable quality excellence. The distinction between learning organizations and unlearning organizations may prove as significant as earlier distinctions between learning organizations and non-learning organizations.

Practical Implications

Several From a practical perspective, this review provides quality managers and organizational leaders with actionable frameworks and insights for implementing unlearning initiatives. The five-dimensional taxonomy enables practitioners to diagnose their specific unlearning needs by identifying relevant triggers, determining primary levels requiring attention, selecting appropriate mechanisms, clarifying desired outcomes, and recognizing contextual factors affecting implementation. Rather than generic change management approaches, the taxonomy supports context-specific unlearning strategies tailored to particular organizational situations.

The review identifies several practical recommendations for quality managers. First, organizations should implement regular knowledge obsolescence assessments evaluating whether existing quality practices, procedures, and assumptions remain relevant and effective. These assessments should employ explicit criteria including regulatory compliance, effectiveness in achieving intended results, efficiency compared to alternative approaches, and strategic alignment with organizational objectives. Without systematic assessment processes, obsolete knowledge accumulates indefinitely because organizational inertia favors retention over elimination.

Second, organizations should establish explicit unlearning mechanisms integrated into quality management systems rather than treating unlearning as ad hoc response to crisis. These mechanisms should include sunset provisions requiring periodic relevance reviews for all documented procedures, version control systems enabling archiving rather than permanent retention, delegated authority structures enabling unlearning decisions without requiring senior approval for routine obsolescence, and knowledge audit processes systematically identifying unlearning candidates. By institutionalizing unlearning alongside learning, organizations transform unlearning from reactive crisis response to proactive capability.

Third, organizations should provide transition support for individuals and teams experiencing expertise obsolescence. Unlearning creates psychological challenges when professional identities are tied to knowledge becoming obsolete. Support should include retraining programs developing new capabilities, role redesign creating positions leveraging transferable skills, career development planning identifying future-oriented development paths, and psychological safety enabling acknowledgment of uncertainty during transitions. Organizations that treat unlearning as individual deficiency rather than organizational development opportunity create resistance that undermines unlearning initiatives.

Fourth, organizations should balance unlearning with learning through integrated approaches addressing both knowledge elimination and knowledge acquisition. Training programs should explicitly identify obsolete practices being replaced, helping participants clear cognitive space for new knowledge. Quality system updates should simultaneously eliminate outdated components and introduce improvements rather than merely layering new atop old. Change initiatives should articulate what must be unlearned as clearly as what must be learned, providing explicit guidance regarding practices requiring abandonment.

Fifth, organizations should leverage technology to support unlearning through knowledge management systems with sunset capabilities, analytics identifying obsolete procedures, and collaboration platforms facilitating knowledge consolidation. Technology can automate certain unlearning aspects such as removing outdated documents from active systems while preserving historical archives, flagging procedures never accessed as unlearning candidates, and enabling virtual collaboration for knowledge consolidation initiatives. However,

technology proves insufficient without corresponding attention to social, political, and psychological dimensions of unlearning.

Limitations

The review also identifies several important limitations constraining interpretation and generalization of findings. First, the restriction to English-language articles in four databases potentially excluded relevant research published in other languages or outlets, particularly research from Asian, Latin American, and other non-Western contexts that may offer distinctive perspectives on unlearning in quality management. Second, publication bias likely affects the sample, as academic journals preferentially publish positive findings over null results, potentially inflating perceived benefits of unlearning while obscuring contexts where unlearning proves ineffective or counterproductive. Third, conceptual heterogeneity complicates synthesis, as authors employ varied definitions and operationalizations of unlearning ranging from passive forgetting to active elimination to cognitive reframing. This definitional diversity enables rich theoretical development but hinders precise integration across studies. Fourth, methodological limitations include predominance of conceptual and qualitative studies limiting generalizability, scarcity of longitudinal research constraining understanding of process dynamics, and absence of validated measurement instruments preventing rigorous hypothesis testing.

Future Research Agenda

Based on identified gaps, the review proposes a comprehensive future research agenda organized around six themes. First, measurement and operationalization research should develop and validate instruments assessing unlearning at individual, team, and organizational levels within quality contexts. Researchers need reliable measures distinguishing unlearning from non-use or temporary suspension, indicators signaling unlearning completion versus superficial adoption, and quantitative approaches assessing unlearning intensity, speed, and thoroughness. Methodological approaches should include scale development studies employing psychometric validation, longitudinal designs tracking knowledge disappearance markers, multi-rater assessments capturing different stakeholder perspectives, and computational text analysis identifying unlearning signals in organizational communication.

Second, research examining unlearning-performance relationships should systematically test hypothesized linkages between unlearning and various quality performance indicators. Specific questions include: Under what conditions does unlearning improve versus harm quality performance? What mediating mechanisms link unlearning to quality outcomes? How do temporal dynamics affect unlearning-performance relationships? What optimal unlearning levels balance obsolescence elimination and capability preservation? Methodological approaches should include longitudinal studies with repeated quality performance measurement before, during, and after unlearning initiatives, quasi-experimental designs comparing unlearning and non-unlearning organizations, structural equation modeling testing mediation hypotheses, and regression discontinuity designs exploiting natural unlearning triggers.

Third, research on contextual contingencies should identify boundary conditions specifying when and where specific unlearning approaches work best. Questions include: How do national cultures affect unlearning processes and effectiveness? Do unlearning requirements differ across manufacturing, service, and knowledge industries? How does organizational size, age, or structure moderate unlearning capability? What role do industry maturity, regulatory intensity, or competitive dynamics play? Methodological approaches should include comparative case studies across contexts, meta-analyses testing context as moderator, configurational approaches such as fuzzy-set qualitative comparative analysis identifying context-specific unlearning recipes, and cross-cultural studies examining cultural value impacts.

Fourth, research on process dynamics and micro-foundations should examine actual unlearning processes including how knowledge disappears, what psychological and social mechanisms operate, and what implementation steps prove necessary. Questions include: What specific activities constitute effective unlearning processes in quality contexts? How do individuals experience unlearning psychologically and emotionally? What social dynamics facilitate or inhibit team-level unlearning? How do unlearning processes unfold temporally with identifiable stages? Methodological approaches should include ethnographic studies observing unlearning initiatives in real-time, experience sampling methods capturing moment-to-moment unlearning experiences, social network analysis tracking knowledge flow changes, and process tracing methods mapping unlearning activity sequences.

Fifth, research on technology's role should investigate how technologies enable or constrain unlearning while simultaneously requiring unlearning themselves. Questions include: How can artificial intelligence support identification of obsolete quality knowledge? What roles can knowledge management systems play in facilitating

systematic unlearning? How does technology-embedded knowledge resist unlearning efforts? Can simulation technologies enable risk-free unlearning experimentation? Methodological approaches should include design science research developing unlearning-support technologies, field experiments testing technology-enabled unlearning interventions, computational modeling simulating unlearning dynamics, and qualitative studies examining human-technology interaction during unlearning.

Sixth, research on unlearning capability development should examine how unlearning capability develops and can be deliberately cultivated. Questions include: What organizational practices build unlearning capability over time? How do organizations learn to unlearn more effectively? Can unlearning capability constitute sustainable competitive advantage? What training approaches develop individual unlearning competence? Methodological approaches should include longitudinal studies tracking capability development, action research implementing capability-building interventions, surveys examining capability antecedents and consequences, and qualitative studies investigating capability-building practices.

Concluding Remarks

In conclusion, organizational unlearning represents a critical yet frequently overlooked dimension of quality management and continuous improvement. As business environments accelerate their rate of change driven by digital transformation, sustainability imperatives, and evolving stakeholder expectations, the ability to unlearn obsolete knowledge, practices, and assumptions becomes increasingly essential for quality excellence. Organizations clinging to outdated quality approaches, however successful historically, risk competitive obsolescence and diminished relevance. The journey toward truly learning organizations requires acknowledging that learning alone proves insufficient. Organizations must also become unlearning organizations capable of strategic knowledge discarding, comfortable with productive forgetting, and skilled at continuous renewal.

Quality management, with its systematic approaches, improvement orientation, and evidence-based philosophy, stands well-positioned to lead organizational unlearning capability development. The principles underlying quality management such as customer focus, process thinking, data-driven decision-making, and continuous improvement remain fundamentally valid even as specific applications evolve. Quality professionals understand that improvement requires both addition and subtraction, both learning and unlearning, both preserving what works and discarding what no longer serves organizational objectives. By explicitly integrating unlearning into quality management frameworks, organizations can achieve the dynamic stability required for sustained excellence in increasingly volatile environments.

This review contributes to that integration by organizing existing knowledge, proposing a comprehensive taxonomy, and charting paths toward deeper understanding. The five-dimensional framework spanning triggers, levels, mechanisms, outcomes, and contexts provides structure for both research and practice. However, significant work remains to transform unlearning from conceptual possibility to systematic organizational capability. Future research answering the questions posed in this review will further enhance both theoretical understanding and practical application of unlearning in quality management contexts.

Quality management has evolved from inspection to prevention, from quality control to quality assurance to total quality management, from reactive problem-solving to proactive improvement. The next evolution may involve integrating unlearning capability, recognizing that organizational memory requires not only accumulation but also strategic discarding, not only learning but also deliberate forgetting, not only preserving best practices but also abandoning practices that were once best but no longer serve organizational objectives. As organizations navigate digital transformation, sustainability imperatives, and unprecedented market volatility, the ability to unlearn may distinguish quality leaders from quality laggards. This review contributes to building that distinguishing capability.

References

Akgün, A. E., Byrne, J. C., Lynn, G. S., & Keskin, H. (2007). Organizational unlearning as changes in beliefs and routines in organizations. *Journal of organizational change management*, *20*(6), 794-812.

Akgün, A. E., Lynn, G. S., & Byrne, J. C. (2006). Antecedents and consequences of unlearning in new product development teams. *Journal of Product Innovation Management*, *23*(1), 73-88.

Antony, J., Krishan, N., Cullen, D., & Kumar, M. (2012). Lean Six Sigma for higher education institutions (HEIs) Challenges, barriers, success factors, tools/techniques. *International Journal of Productivity and Performance Management*, *61*(8), 940-948.

Antony, J., Sony, M., McDermott, O., Jayaraman, R., & Flynn, D. (2023). An exploration of organizational readiness factors for Quality 4.0: an intercontinental study and future research directions. *International Journal of Quality & Reliability Management*, *40*(2), 582-606.

Antonacopoulou, E. P. (2009). Impact and scholarship: Unlearning and practising to co-create actionable knowledge. *Management learning*, *40*(4), 421-430.
Becker, K. (2005). Individual and organisational unlearning: directions for future research. *International journal of organisational behaviour*, *9*(7), 659-670.
Bessant, J., Caffyn, S., & Gallagher, M. (2001). An evolutionary model of continuous improvement behaviour. *Technovation*, *21*(2), 67-77.
Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, *3*(2), 77-101.
Caldwell, K., Henshaw, L., & Taylor, G. (2011). Developing a framework for critiquing health research: an early evaluation. *Nurse education today*, *31*(8), e1-e7.
Campanelli, A. S., & Parreiras, F. S. (2015). Agile methods tailoring–A systematic literature review. *Journal of Systems and Software*, *110*, 85-100.
Cegarra-Navarro, J. G., & Dewhurst, F. W. (2006). Linking shared organisational context and relational capital through unlearning: An initial empirical investigation in SMEs. *The learning organization*, *13*(1), 49-62.
Gabriel Cegarra-Navarro, J., Sánchez-Vidal, M. E., & Cegarra-Leiva, D. (2011). Balancing exploration and exploitation of knowledge through an unlearning context: An empirical investigation in SMEs. *Management Decision*, *49*(7), 1099-1119.
Cegarra-Navarro, J. G., & Wensley, A. (2019). Promoting intentional unlearning through an unlearning cycle. Journal of Organizational Change Management, 32(1), 67-79.
Cegarra-Navarro, J. G., Eldridge, S., & Wensley, A. K. (2014). Counter-knowledge and realised absorptive capacity. *European Management Journal*, *32*(2), 165-176.
Dean Jr, J. W., & Bowen, D. E. (1994). Management theory and total quality: improving research and practice through theory development. *Academy of management review*, *19*(3), 392-418.
Holan, P. M. D., & Phillips, N. (2004). Remembrance of things past? The dynamics of organizational forgetting. *Management science*, *50*(11), 1603-1613.
Denyer, D., & Tranfield, D. (2009). Producing a systematic review.
Easterby-Smith, M., & Lyles, M. A. (2011). In praise of organizational forgetting. *Journal of Management Inquiry*, *20*(3), 311-316.
Fernandez, V., & Sune, A. (2009). Organizational forgetting and its causes: an empirical research. *Journal of organizational change management*, *22*(6), 620-634.
Fiol, M., & O'Connor, E. (2017). Unlearning established organizational routines–part I. *The Learning Organization*, *24*(1), 13-29.
Fonseca, L. (2015). From Quality Gurus and TQM to ISO 9001: 2015: a review of several quality paths.
García-Morales, V. J., Llorens-Montes, F. J., & Verdú-Jover, A. J. (2006). Antecedents and consequences of organizational innovation and organizational learning in entrepreneurship. *Industrial Management & Data Systems*, *106*(1), 21-42.
Govindan, K., & Hasanagic, M. (2018). A systematic review on drivers, barriers, and practices towards circular economy: a supply chain perspective. *International journal of production research*, *56*(1-2), 278-311.
Grisold, T., Mendling, J., Otto, M., & vom Brocke, J. (2021). Adoption, use and management of process mining in practice. *Business Process Management Journal*, *27*(2), 369-387.
Hatch, N. W., & Dyer, J. H. (2004). Human capital and learning as a source of sustainable competitive advantage. *Strategic management journal*, *25*(12), 1155-1178.
Bo, H. (1981). How organizations learn and unlearn. *Handbook of organizational design (1)*, 3-27.
Hislop, D., Bosley, S., Coombs, C. R., & Holland, J. (2014). The process of individual unlearning: A neglected topic in an under-researched field. *Management Learning*, *45*(5), 540-560.
Juran, J. M. (1999). *Juran's quality handbook*.
Klammer, A., & Gueldenberg, S. (2019). Unlearning and forgetting in organizations: a systematic review of literature. *Journal of Knowledge management*, *23*(5), 860-888.
Laureani, A., & Antony, J. (2019). Leadership and Lean Six Sigma: a systematic literature review. *Total Quality Management & Business Excellence*, *30*(1-2), 53-81.
Leonard-Barton, D. (1992). Core capabilities and core rigidities: A paradox in managing new product development. *Strategic management journal*, *13*(S1), 111-125.
Martin de Holan, P., Phillips, N., & Lawrence, T. B. (2004). Managing organizational forgetting. *MIT Sloan Management Review*, 45(2), 45-51.
Maull, R., Brown, P., & Cliffe, R. (2001). Organisational culture and quality improvement. *International Journal of Operations & Production Management*, *21*(3), 302-326.
Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Bmj*, *339*.
Nystrom, P., & Starbuck, W. H. (2015). To avoid organizational crises, unlearn. *Unlearn (December 25, 2015)*.

Petticrew, M., & Roberts, H. (2008). *Systematic reviews in the social sciences: A practical guide*. John Wiley & Sons.
Prajogo, D. I., & Sohal, A. S. (2003). The relationship between TQM practices, quality performance, and innovation performance: An empirical examination. *International journal of quality & reliability management*, *20*(8), 901-918.
Rushmer, R., & Davies, H. T. O. (2004). Unlearning in health care. *BMJ Quality & Safety*, *13*(suppl 2), ii10-ii15.
Sampaio, P., Saraiva, P., & Guimarães Rodrigues, A. (2009). ISO 9001 certification research: questions, answers and approaches. *International journal of quality & reliability management*, *26*(1), 38-58.
Santos-Vijande, M. L., & Álvarez-González, L. I. (2007). Innovativeness and organizational innovation in total quality oriented firms: The moderating role of market turbulence. *Technovation*, *27*(9), 514-532.
Sarkis, J., Zhu, Q., & Lai, K. H. (2011). An organizational theoretic review of green supply chain management literature. *International journal of production economics*, *130*(1), 1-15.
Shipton, H. (2006). Cohesion or confusion? Towards a typology for organizational learning research 1. *International Journal of Management Reviews*, *8*(4), 233-252.
Sinkula, J. M. (2002). Market-based success, organizational routines, and unlearning. *Journal of Business & Industrial Marketing*, *17*(4), 253-269.
Sony, M., & Naik, S. (2020). Industry 4.0 integration with socio-technical systems theory: A systematic review and proposed theoretical model. *Technology in society*, *61*, 101248.
Tarí, J. J., & Dick, G. (2016). Trends in quality management research in higher education institutions. *Journal of Service Theory and Practice*, *26*(3).
Tortorella, G. L., Giglio, R., & Van Dun, D. H. (2019). Industry 4.0 adoption as a moderator of the impact of lean production practices on operational performance improvement. *International journal of operations & production management*, *39*(6/7/8), 860-886.
Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British journal of management*, *14*(3), 207-222.
Tsai, F. S., Hsieh, L. H., Fang, S. C., & Lin, J. L. (2009). The co-evolution of business incubation and national innovation systems in Taiwan. *Technological Forecasting and Social Change*, *76*(5), 629-643.
Tsang, E. W. (2008). Transferring knowledge to acquisition joint ventures: an organizational unlearning perspective. *Management Learning*, *39*(1), 5-20.
Tsang, E. W., & Zahra, S. A. (2008). Organizational unlearning. *Human relations*, *61*(10), 1435-1462.
Zahra, S. A., Abdelgawad, S. G., & Tsang, E. W. (2011). Emerging multinationals venturing into developed economies: Implications for learning, unlearning, and entrepreneurial capability. *Journal of management inquiry*, *20*(3), 323-330.
Zeithaml, V. A., Jaworski, B. J., Kohli, A. K., Tuli, K. R., Ulaga, W., & Zaltman, G. (2020). A theories-in-use approach to building marketing theory. *Journal of Marketing*, *84*(1), 32-51.
Zhao, Y., Li, Y., Lee, S. H., & Bo Chen, L. (2011). Entrepreneurial orientation, organizational learning, and performance: Evidence from China. *Entrepreneurship theory and practice*, *35*(2), 293-317.
Zu, X., Fredendall, L. D., & Douglas, T. J. (2008). The evolving theory of quality management: the role of Six Sigma. *Journal of operations Management*, *26*(5), 630-650.