Uluslararası Psiko-Sosyal Eğitim Araştırmaları Dergisi
Cilt 6, Sayı 10, 2026, 97-121
Gönderim: 14/10/2025 **Kabul**: 26/02/2026

e-ISSN: 2822-4574
Araştırma Makalesi
https://doi.org/10.5281/zenodo.21174254

Trends In Argumentative Thinking Skills: A Systematic Review

Derya BEKİROĞLU[1], Nur ÜTKÜR GÜLLÜHAN[2]

Abstract

Argumentative thinking skills play a fundamental role in developing critical reasoning, evidence-based decision-making, and constructive discourse. This systematic review examines research trends on argumentative thinking skills published between 2014 and 2024, aiming to reveal how the field has evolved regarding publication trends, methodological approaches, participant groups, and thematic orientations. A total of 67 peer-reviewed studies indexed in the Web of Science (WoS) database were analyzed in line with the PRISMA 2020 framework. Data were collected through systematic document review, and coding reliability was verified by expert opinion. The findings indicate that most studies focus on argumentative writing, critical thinking, and discourse-based learning, with university students being the predominant participant group. Quantitative research designs, convenience sampling, scale-based measurement tools, and statistical analyses such as t-test and ANOVA were identified as the most frequently used methods. Recently, a significant increase in integrating digital technologies and artificial intelligence has been observed. However, the results highlight critical gaps, including a shortage of longitudinal research, a lack of studies focusing on early childhood and primary education, and a scarcity of culturally responsive approaches. By synthesizing these trends and gaps, this study provides a crucial roadmap for future research to develop more inclusive, technology-integrated, and methodologically diverse pedagogical models.

Keywords: Argumentative thinking skills, Systematic review, PRISMA, Educational trends, Educational technologies.

Tartışmacı Düşünme Becerilerindeki Eğilimler: Sistematik Bir Derleme

Özet

Tartışmacı düşünme becerileri; eleştirel akıl yürütme, kanıta dayalı karar verme ve yapıcı söylem süreçlerinin geliştirilmesinde temel bir rol oynamaktadır. Bu sistematik derleme, 2014–2024 yılları arasında yayımlanan tartışmacı düşünme becerileri üzerine yapılan araştırmaları; yayın eğilimleri, yöntemsel yaklaşımlar, katılımcı grupları ve tematik yönelimler açısından incelemeyi amaçlamaktadır. Web of Science (WoS) veri tabanında indekslenen toplam 67 hakemli çalışma, PRISMA 2020 çerçevesi doğrultusunda analiz edilmiştir. Veriler sistematik belge incelemesi yoluyla toplanmış ve kodlama güvenirliği uzman görüşü ile doğrulanmıştır. Bulgular, çalışmaların çoğunun tartışmacı yazma, eleştirel düşünme ve söyleme dayalı öğrenmeye odaklandığını, katılımcı grubunun ise ağırlıklı olarak üniversite öğrencilerinden oluştuğunu göstermektedir. Nicel araştırma desenleri, kolay ulaşılabilir durum örneklemesi, ölçek temelli ölçme araçları ile t-testi ve ANOVA gibi istatistiksel analizlerin en sık kullanılan yöntemler olduğu belirlenmiştir. Son yıllarda, dijital teknolojilerin ve yapay zekânın sürece entegrasyonunda belirgin bir artış gözlenmiştir. Bununla birlikte; uzunlamasına araştırmaların sınırlı olduğu, erken çocukluk ve ilkokul dönemine odaklanan çalışmaların yetersiz kaldığı ve kültürel olarak duyarlı yaklaşımların eksik olduğu tespit edilmiştir. Bu çalışma, mevcut literatürdeki eğilimleri ve boşlukları sentezleyerek gelecekteki araştırmaların daha kapsayıcı, teknoloji entegreli ve yöntem çeşitliliğine sahip pedagojik modeller geliştirmesine yönelik kritik bir yol haritası sunmaktadır.

Anahtar Kelimeler: Tartışmacı düşünme becerileri, Sistematik derleme, PRISMA, Eğitim eğilimleri, Eğitim teknolojileri.

[1] PhD Candidate, Istanbul University-Cerrahpaşa, Graduate School of Education, drybkrgl@gmail.com
[2] Assoc.Prof. Dr., Istanbul University- Cerrahpaşa, Hasan Âli Yücel Faculty of Education, nur.utkur@iuc.edu.tr

Extended Abstract

Giriş: Tartışmacı düşünme, bireylerin bir konudaki fikirlerini mantıksal bir yapıda düzenledikleri, neden-sonuç ilişkileri kurdukları ve kanıtlar ile sağlam gerekçelerle desteklenen karşı argümanlar geliştirdikleri süreci ifade eder; bu beceri yalnızca kişinin kendi bakış açısını ifade etmesini değil, aynı zamanda alternatif görüşleri de eleştirel olarak değerlendirmesini sağlar (Kuhn, 1992; Osborne, 2010). Eğitim bağlamında tartışmacı düşünme; eleştirel ve analitik düşünmeyi teşvik etmede, kavramsal anlamayı derinleştirmede ve iletişim becerilerini geliştirmede merkezi bir rol oynamaktadır (Erduran & Jiménez-Aleixandre, 2008; Driver vd., 2000). Tarihsel kökenleri Antik Yunan'da, Sokrates'in daha derin yansımaları teşvik etmek için sorgulama yöntemini kullanmasına ve Aristoteles'in Rhetoric ile Organon gibi eserlerde mantıksal akıl yürütme ilkelerini sistemleştirmesine kadar uzanmaktadır (Kienpointner, 2020; Schwarz & Baker, 2017; Tindale, 2004). Modern dönemde ise Toulmin'in (1958) iddia, veri, gerekçe ve çürütmeden oluşan argümantasyon modeli, tartışma yapılarının sistematik bir analizine olanak tanımıştır (Tippett, 2009). 21. yüzyılda tartışmacı düşünmenin kapsamı, dil ve fen eğitiminden; tartışmacı düşünme öğrenme ortamları için yapay zekâ destekli yardımlara, disiplinler arası STEM uygulamalarına ve çevrimiçi tartışma platformlarına kadar genişlemiştir (Noroozi vd., 2023; García-Mila vd., 2023). Literatür incelendiğinde, mevcut çalışmaların genellikle belirli disiplinlere odaklandığı, sınırlı yaş gruplarını (genellikle ortaöğretim veya yükseköğretim düzeyi) hedef aldığı ve kısa vadeli araştırma desenleriyle yürütüldüğü görülmektedir. Önceki derleme çalışmaları yapay zekâ entegrasyonu, metodolojik çeşitlilik gibi güncel eğilimleri haritalandırmakta yetersiz kalmıştır (Erduran vd., 2015; Bekiroğlu & Ütkür-Güllühan, 2022; Samson & Blanchard, 2012). Bu çalışma, alandaki tematik ve yöntemsel yönelimleri kapsamlı bir şekilde sentezleyerek literatürdeki bu önemli boşluğu doldurmaktadır.

Yöntem: Bu çalışmada, belirli bir araştırma sorusuna yönelik mevcut literatürü kapsamlı, titiz ve nesnel bir şekilde sentezlemek amacıyla sistematik derleme yöntemi benimsenmiştir. Araştırma süreci, metodolojik şeffaflık ve sürdürülebilirlik sağlamak adına PRISMA 2020 çerçevesi doğrultusunda yürütülmüştür. Veri toplama sürecinde doğrudan Web of Science veri tabanında indekslenen ve başlığında ya da anahtar kelimelerinde doğrudan "argumentative thinking" veya "argumentative thinking skills" ifadesi geçen, İngilizce veya Türkçe dillerinde tam metnine ulaşılabilen çalışmalar dahil edilme kriteri olarak belirlenmiştir. Yapılan ilk taramada 134 kayda ulaşılmış; 4 mükerrer kayıt çıkarıldıktan sonra 130 çalışma başlık ve özet taramasına alınmıştır. Bu aşamada kriterleri karşılamayan 37 çalışma elenmiş, kalan 93 tam metinden 15'ine erişilemediği için 78 çalışma uygunluk değerlendirmesine tabi tutulmuştur. Son tahlilde, hakemli olmayan (5 makale), odağı tartışma becerisi olmayan (3 makale) ve dil kriterine uymayan (3 makale) çalışmalar da dışarıda bırakılarak toplam 67 hakemli makale analize dahil edilmiştir. Dahil edilen çalışmaların metodolojik kalitesi, eğitim araştırmalarına uyarlanmış Newcastle-Ottawa Ölçeği kullanılarak iki bağımsız hakem tarafından değerlendirilmiş ve hakemler arası uzlaşı düzeyinin yüksek olduğu saptanmıştır. Elde edilen veriler betimsel analiz tekniğiyle çözümlenerek frekans ve yüzde dağılımları halinde tablolara dönüştürülmüştür. Çalışmalar kodlanırken iki ana alt tema oluşturulmuştur: Birinci alt tema "Eleştirel Düşünme, Yapay Zekâ ve Matematik Becerileri" ; ikinci alt tema ise "Tartışmacı Yazma ve Okuma Becerileri" olarak yapılandırılmıştır.

Bulgular: 1. Eleştirel Düşünme, Yapay Zekâ ve Matematik Temalı Çalışmalar: Bu alt temadaki çalışmaların yıllara göre dağılımı incelendiğinde, özellikle 2020 sonrasında belirgin bir artış yaşandığı ve en yüksek yayına 2021 yılında (f=6) ulaşıldığı görülmektedir. Çalışmaların en sık yayımlandığı dergilerin başında Thinking Skills and Creativity (f=6) gelmektedir. Metodolojik açıdan incelendiğinde, nicel deneysel desenler (f=11) ile karma yöntem eş zamanlı açıklayıcı desenlerin (toplam karma f=11, açıklayıcı sıralı f=7) sahaya hâkim olduğu, saf nitel araştırmaların (f=2) ise oldukça sınırlı kaldığı tespit edilmiştir. Örnekleme yöntemlerinde kolay ulaşılabilirlik ve amaca yönelik örneklemeler (f=23) baskındır. Katılımcı düzeyinde ise üniversite öğrencileri (f=10) ve lise öğrencileri (f=9) ağırlıktayken, ilkokul seviyesi (f=2) oldukça sınırlıdır. Örneklem büyüklüğü olarak araştırmacıların sınıf tabanlı müdahalelerin doğası gereği 31-100 kişi aralığını (f=15) tercih ettikleri görülmektedir. Veri toplama aracı olarak ölçekler (f=13) ilk sıradadır. İstatistiksel analizlerde ise betimsel istatistiklerin yanı sıra çıkarımsal istatistik olarak grup karşılaştırmalarına imkân tanıyan ANOVA/ANCOVA (f=13) ve t-testi (f=9) en sık başvurulan yöntemler olmuştur. 2. Tartışmacı Yazma ve Dil Becerileri Temalı Çalışmalar: Bu gruptaki akademik üretimin 2014-2019 yılları arasında durağan seyrettiği (yıllık f=2-3), ancak 2020 yılından itibaren büyük bir ivme kazanarak 2021 (f=8) ve 2023 (f=9) yıllarında zirve noktalarına ulaştığı saptanmıştır. Yayınların dağılımında RED- Revista de Educacion a Distancia (f=8) ve Asian-Pacific Journal of Second and Foreign Language Education (f=7) öne çıkan mecralardır. Metodolojik mimaride nicel yaklaşımlar (f=21) ve deneysel tasarımlar (f=16) başı çekmekte, bunları gömülü desenlerin ağırlıkta olduğu karma yöntemler (f=17) izlemektedir. Katılımcıların seçiminde yine olasılıksız örnekleme yöntemleri (kolay ulaşılabilir f=31, amaçlı f=12) mutlak bir egemenliğe sahiptir. Bu temada da örneklemlerin ağırlıklı olarak üniversite (f=19) ve lise (f=12) seviyesinde yoğunlaştığı belirlenmiştir. Çalışmaların büyük bir kısmında 31-100 kişilik sınıf ölçekli gruplar (f=26) örneklem büyüklüğü olarak seçilmiştir. Veri toplama aracında ölçekler (f=36) ve veri çeşitlendirmesi sağlamak adına yarı yapılandırılmış görüşmeler (f=17) yoğunluktadır. Analiz süreçlerinde sürekli verilerin merkezi eğilimini göstermek adına aritmetik ortalama (f=23) ön plandadır; çıkarımsal testlerde ise deneysel grupların başarı farklarını ölçen t-testi (f=18) ile ANOVA/ANCOVA (f=11) yoğun şekilde kullanılmıştır.

Tartışma ve Sonuç: Sistematik derlemeden elde edilen sonuçlar, tartışmacı düşünme becerileri alanındaki literatürün son on yılda özellikle dijital teknolojiler, yapay zekâ araçları ve matematik entegrasyonu ekseninde dinamik bir gelişim gösterdiğini ortaya koymaktadır. Yayınların 2021 ve 2023 yıllarında kümelenmesi, pandeminin hızlandırdığı dijital dönüşüm ve uzaktan eğitim modelleriyle doğrudan ilişkilidir. Ancak bu tematik zenginleşmeye karşılık, alanda metodolojik bir durağanlığın yaşandığı görülmektedir. Araştırmaların çok büyük bir kısmı kısa vadeli, yarı deneysel ve sadece iki grup arasındaki farkı test eden tek değişkenli istatistiksel modellere sıkışmış durumdadır. Sentez sonuçları alandaki üç büyük yöntemsel ve tematik boşluğa dikkat çekmektedir: İlk olarak, katılım kaybı ve müfredat parçalanması gibi yapısal engeller nedeniyle uzunlamasına araştırmalar yok denecek kadar azdır. İkinci olarak, tartışmacı düşünme üst düzey bir bilişsel süreç olarak görüldüğünden örneklemler yetişkin ve ergen gruplarına kaymış, erken çocukluk ve ilkokul dönemindeki çocukların bu beceriyi nasıl kazandığına dair çalışmalar yetersiz kalmıştır. Son olarak, çok değişkenli dinamik ilişkileri açıklayabilecek istatistiksel modeller ile doğal öğrenme ortamlarındaki argüman örüntülerini yakalayacak söylem/içerik analizleri arka planda kalmıştır. Gelecekteki araştırmalar için bu çalışma iki temel yol haritası sunmaktadır: Tematik olarak, teknolojiyi sadece bir araç olarak görmeyi bırakıp, öğrencilerin yapay zekâ tarafından üretilen taraflı veya hatalı bilgileri eleştirel olarak değerlendirebilecekleri insan-yapay zekâ ortaklaşa tartışma modellerine ve algoritmik okuryazarlığa odaklanılmalıdır. Metodolojik olarak ise, öğrenme analitikleri, göz izleme gibi gelişmiş veri yaklaşımlarından yararlanılmalı, kültürel olarak duyarlı modeller geliştirilmeli ve becerilerin kalıcılığını test edecek uzun vadeli boylamsal desenler pedagojik modellere entegre edilmelidir.

INTRODUCTION

Argumentative thinking refers to the process by which individuals organise their ideas on a topic in a logical structure, establish cause–effect relationships, and develop counterarguments supported by evidence and sound reasoning. This skill enables not only the articulation of one's own perspective but also the critical evaluation of alternative viewpoints (Kuhn, 1992; Osborne, 2010). While argumentation broadly refers to the exchange of claims and evidence in dialogue, argumentative thinking skills focus more specifically on the cognitive processes involved in constructing and evaluating arguments. In educational contexts, argumentative thinking plays a central role in promoting critical and analytical thinking, deepening conceptual understanding, and enhancing communication skills (Erduran & Jiménez-Aleixandre, 2008; Driver et al., 2000).

Historically, the roots of argumentation can be traced back to Ancient Greece, where Socrates employed the method of questioning to encourage deeper reflection, and Aristotle systematised the principles of logical reasoning in works such as Rhetoric and Organon (Kienpointner, 2020; Schwarz & Baker, 2017; Tindale, 2004). In the modern era, Toulmin's (1958) argumentation model comprising claim, data, warrant, and rebuttal enabled a systematic analysis of argumentative structures (Tippett, 2009). In the 21st century, the scope of argumentative thinking has extended language and science education to areas such AI-enhanced support for argumentative thinking learning environments, interdisciplinary STEM practices, and online discussion platforms (Noroozi et al., 2023; García-Mila et al., 2023)

Significance and Rationale

In contemporary education systems, higher-order thinking skills such as critical thinking, digital literacy, and interdisciplinary problem-solving are key priorities. Argumentation skills are central to achieving these goals (Andriessen, 2006; Osborne et al., 2004). However, a review of the literature reveals that existing studies often focus on specific disciplines (e.g., science education), limited age groups (typically secondary or tertiary level), and short-term research designs.

Previous review and meta-synthesis studies (Erduran et al., 2015; Bekiroğlu & Ütkür-Güllühan, 2022; Samson & Blanchard, 2012) have highlighted the importance of argumentation but have not provided a systematic mapping of thematic trends, methodological diversity, and emerging areas such as AI-enhanced argumentation or interdisciplinary applications. Moreover, there is a lack of comprehensive synthesis addressing both thematic orientations and methodological quality indicators in this field. Therefore, this study addresses a significant gap by offering a comprehensive synthesis of thematic and methodological orientations in the field.

Purpose of the Study

In line with this rational the purpose of this study is to identify research trends in argumentative thinking skills by analyzing articles published in the Web of Science (WoS) database between 2014 and 2024. The study analyzes these trends through specific thematic domains coded as Mathematics (AMATH), Technology-AI (ATECH), Writing and Language (AWRL), and Critical Thinking (ACT). Specifically, the study addresses the following research questions corresponding to the main thematic findings:

1. What are the research trends regarding the themes of Mathematics (AMATH) and Technology-AI (ATECH) in terms of in terms of educational contexts, content-based issues, distribution by year and journal, as well as methodological characteristics (research design, sample, and data analysis)?

2. What are the research trends regarding the themes of Writing and Language (AWRL) and Critical Thinking (ACT) in terms of educational contexts, content-based issues, distribution by year and journal, as well as methodological characteristics (research design, sample, and data analysis)?

METHOD

Research Design

This study employed the systematic review method, which is a comprehensive, rigorous, and objective approach to synthesizing existing literature related to a clearly defined research question. In systematic reviews, all relevant studies meeting predetermined criteria are identified, critically evaluated, and synthesized to reveal overarching trends and gaps in the field (Grant & Booth, 2009).

To ensure methodological transparency, the review process was conducted in accordance with the PRISMA 2020 guidelines (Preferred Reporting Items for Systematic Reviews and Meta-Analyses). The PRISMA flow diagram (Figure 1) illustrates the four key stages of the review: identification, screening, eligibility, and inclusion.

The methodological quality of the included studies was evaluated using the Newcastle–Ottawa Scale (NOS), adapted for non-randomized studies. This scale assesses three dimensions:

1. Selection of the study groups,

2. Comparability of the groups, and

3. Outcome measurement.

Two independent reviewers conducted the NOS evaluation using a star-based rating system, and any disagreements were resolved through discussion until consensus was reached. Quality assessment results were visualized using the Robvis tool, generating both a traffic-light plot and a bar chart for clearer interpretation.

Following the PRISMA guided study selection process, the methodological quality of the 67 included studies was evaluated using adapted criteria from the Newcastle–Ottawa Scale (NOS) for educational research. Two independent reviewers assessed each study across four domains “Study Selection, Assessment Tool Quality, Reference Standards, and Flow & Timing” with discrepancies resolved through consensus.

Sample Group

The study group consisted of 67 peer-reviewed articles published between 2014 and 2024 in journals indexed in the SCI, ESCI, or SCIE categories of the Web of Science (WoS) database. Purposeful sampling was employed to select studies directly relevant to the research scope. It is important to acknowledge that the exclusion criteria applied in this study—specifically limiting the search to the exact term 'argumentative thinking'—may have narrowed the scope of the review. Comparable research employing broader or related terms such as 'discourse,' 'argumentation skills,' or 'reasoning' might have been excluded. However, this exclusion was a deliberate methodological choice. The primary objective was to isolate the specific cognitive construct of 'thinking' within argumentation, rather than analyzing broader linguistic features or general performance skills often associated with 'discourse' studies. Therefore, while this focus ensures high relevance and conceptual consistency regarding the cognitive dimension, it interprets the field through a strictly defined terminological lens.

Inclusion criteria were as follows:

- Publication date between 2014 and 2024.
- Keywords such as *"argumentative"*, *"argumentative thinking skills"* appearing in the study title.
- Published in journals indexed in SCI, ESCI, or SCIE within the WoS database.
- Accessibility of the full text in English or Turkish.

Exclusion criteria included:

- Conference abstracts, book chapters, and non-peer-reviewed reports.

- Studies where argumentation was mentioned but not the primary focus.

Data Collection Process

In this study, data was collected through document review. Article studies were accessed from "Web of Science" databases. A total of 134 studies were accessed in the literature review of the research. A total of 67 studies on argumentative thinking skills were included in the study according to the inclusion and exclusion criteria as a result of reading the abstracts and checking the titles and keywords. The data collection process is shown in prisma chart flow figure 1.

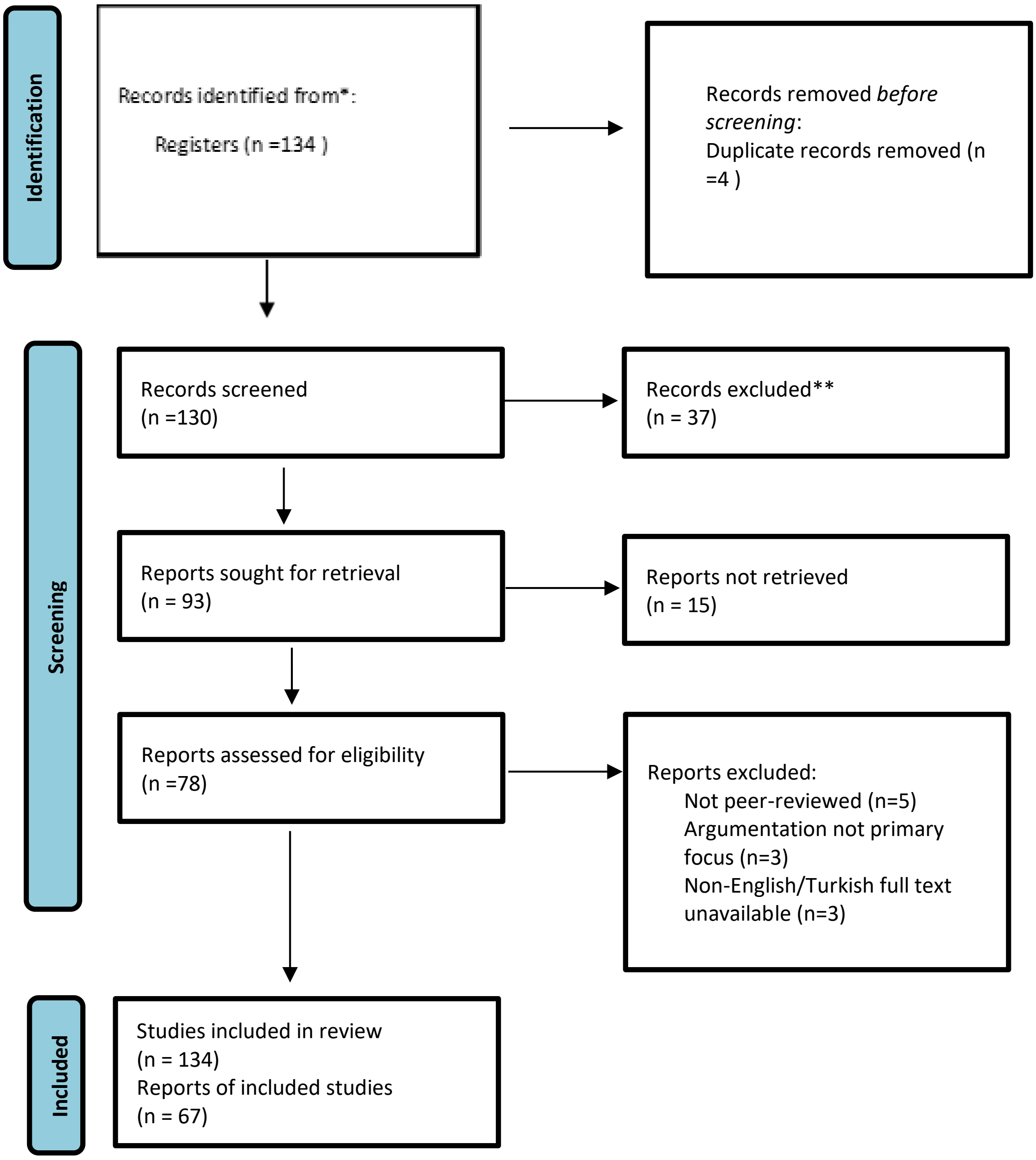

Figure 1. PRISMA Chart

As shown in Figure 1, our "Web of Science" search initially identified 134 records. After removing 4 duplicates, 130 records underwent title and abstract screening, during which 37 were excluded for not meeting the inclusion criteria. Of the remaining 93 full-text articles sought for retrieval, 15 could not be accessed. This left 78 full-text studies assessed for eligibility. In this stage, 11 articles were excluded: 5 were not peer-reviewed, 3 did not have argumentative thinking as their primary focus, and 3 had no full text available in English or Turkish. Consequently, 67 studies met all inclusion criteria and were retained for systematic review. No meta-analysis was performed. This process follows PRISMA 2020 guidelines to ensure methodological transparency and reproducibility.

The methodological quality of the studies is shown in Figure 2 of the Newcastle–Ottawa Scale (NOS).

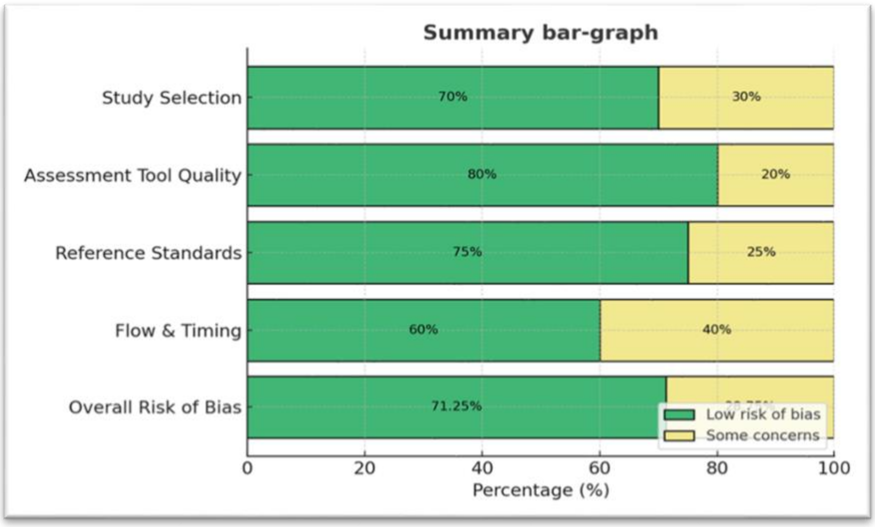

Figure 2. Summary bar-graph

Following the PRISMA-guided study selection process, the methodological quality of the 67 included studies was evaluated using adapted criteria from the Newcastle–Ottawa Scale (NOS) for educational research. Two independent reviewers assessed each study across four domains “Study Selection, Assessment Tool Quality, Reference Standards, and Flow & Timing” with discrepancies resolved through consensus. Two independent reviewers assessed each study within these domains. Any discrepancies in scoring were resolved through discussion and consensus. Inter-rater agreement before consensus was found to be substantial (Cohen’s $\kappa = 0.82$), hich indicates substantial agreement (Landis & Koch, 1977). Any discrepancies were resolved through discussion until full consensus was reached.

As illustrated in Figure 2, in the Study Selection domain, 47 studies (70%) were rated as low risk of bias, while 20 studies (30%) presented some concerns. In the Assessment Tool Quality domain, 54 studies (80%) were judged as low risk, with 13 studies (20%) showing concerns. In the Reference Standards domain, 50 studies (75%) demonstrated a low risk, compared to 17 studies (25%) with some concerns. In the Flow & Timing domain, 40 studies (60%) were assessed as low risk, while 27 studies (40%) exhibited some concerns, often due to incomplete reporting of data collection schedules or follow-up procedures.

Overall, the mean NOS-adapted score across all studies was 7.0 out of 9, reflecting a generally robust methodological quality in the included literature. Nonetheless, clearer participant selection criteria and more detailed reporting of procedural timelines remain areas for improvement in future research.

Coding and Theming Process

A total of 67 studies included in the research were compared and contrasted, and each was coded. As seen in Figure 3, the studies were coded according to the type of skill involved, the discipline the study was related to, and the number of the study.

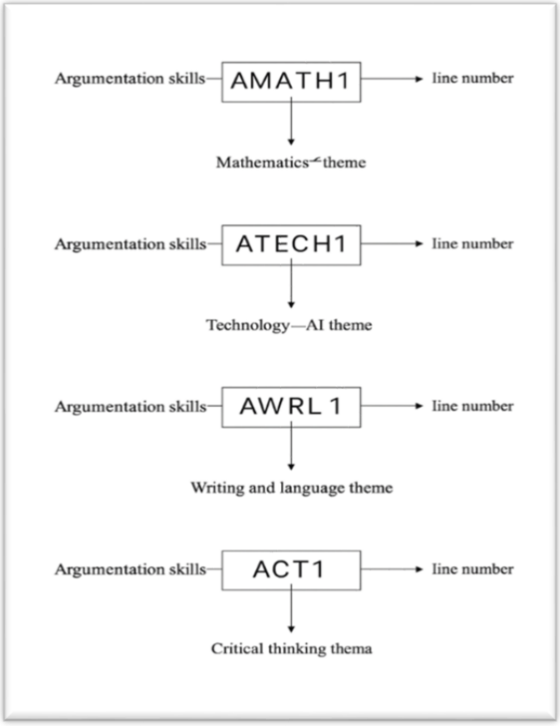

Figure 3. Coding studies

As seen in Figure 3, Studies were coded into four themes: Argumentation–Critical Thinking (ACT),Argumentation–Mathematics (AMATH), Argumentation–Technology and Artificial Intelligence (ATECH), and Argumentative Writing and Language (AWL). After coding the studies, 2 sub themes were created regarding the studies conducted on argumentation thinking skills, as seen in Table 1.

Table 1. Sub-Themes of studies on argumentative thinking skills

	Sub-Themes	n
1	Sub theme of studies conducted on argumentation thinking skills "critical thinking skills / artificial intelligence- chatgpt- mathematical skills"	24
2	Sub theme of studies conducted on argumentative thinking skills "argumentative writing and reading skills"	43
	Total	67

Data Analysis

The data analysis process in this research commenced with a systematic examination of each theme, guided by the parameters outlined in the *Study Review Form* developed by the researchers. This form was designed to ensure that all studies included in the review were assessed consistently across predefined categories, such as research design, sample characteristics, data collection tools, and analytical approaches.

Once the studies were evaluated according to these parameters, the data obtained were organized and prepared for analysis. A descriptive analysis technique was employed to summarize and interpret the collected information, focusing on identifying patterns, trends, and distributions across the selected studies. This approach allowed for the transformation of raw data into meaningful summaries that reflect the methodological and thematic tendencies of the literature.

The results of the descriptive analysis were presented in the form of frequency distributions and tables, providing a clear and structured overview of the findings. These tabular presentations facilitated comparisons between variables and supported the identification of dominant characteristics within the body of research examined. By adopting this methodical approach, the study ensured transparency, replicability, and a comprehensive representation of the data across all reviewed source

FINDINGS

Findings Regarding the Theme of Argumentation Thinking Skills “Critical Thinking Skills/artificial intelligence-mathematics-technology”

The analysis of the studies clustered under the first theme—comprising Critical Thinking (ACT), Mathematics (AMATH), and Technology/AI (ATECH)—reveals significant trends regarding their temporal distribution and thematic evolution.

Distribution Of Studies By Year

Based on the data collected, the distribution of studies conducted on argumentative including critical thinking skills (ACT), artificial intelligence and technology (ATECH), and mathematics (AMATH) by year is shown in Table 2.

Table 2. Distribution of studies by year

Years	Study Code	*f*
2014	ACT13, ACT 21	2
2015	ACT20	1
2016	ACT12	1
2017	ACT14, ACT15	2
2018	ACT16	1
2019	ACT17, ATECH1	2
2020	ACT1, ACT2, ACT3	3
2021	ACT4, ACT5, ACT6, AMATH1, AMATH2, ATECH2	6
2022	ACT12, ACT19	2
2023	ACT7, ACT18	1
2024	ACT9, ACT10, ACT11	3
	Total	24

As shown in Table 2, the number of studies focusing on the theme of ACT, ATECH, and AMATH has increased steadily over the last five years, indicating a growing interest in integrating these domains into argumentative thinking research. The highest concentration was observed in 2021 (f = 6), followed by 2024 (f = 3) and 2020, 2022, and 2023 (each f = 2). In earlier years, such as 2014, 2017, and 2019, the number of studies remained modest (f = 2), while 2015, 2016, and 2018 were represented by only one study each. This distribution suggests a notable increase in research interest after 2020, which may be related to the growing emphasis on digital learning environments, integration of artificial intelligence into education, and the shift towards remote learning during the COVID-19 pandemic.

Distribution of Studies by Journals

Based on the collected data, Table 3 presents the distribution of ACT, ATECH, and AMATH studies across journals.

Table 3. Distribution of studies by journals

Journal	Study Code	f
Jordan Journal Of Modern Languages & Literature	ACT3, ACT8	2
Journal Of Computer Assisted Learnıng	ACT1, ACT,12	2
Arts And Humanıtıes In Hıgher Educatıon	ACT7, ACT11	2
Frontiers In Psychology	ACT10, ACT8, ACT9	3
Thınkıng Skılls And Creatıvıty	ACT2, ACT4, ACT5, ACT20, ACT10, ACT18	6
European Journal Of Psychology Of Educatıon	ACT6, ACT 19	2
Red-Revısta De Educacıon A Dıstancıa	ACT 13, ACT 14, ACT15	3
Internatıonal Journal For Technology In Mathematıcs Educatıon	AMATH1,AMATH2, ATECH1,ATECH2	4
Total		24

As presented in Table 3, the dissemination of research across the ACT, ATECH, and AMATH themes reflects a broadly multidisciplinary landscape rather than relying on a single niche outlet. Thinking Skills and Creativity emerged as the most prominent venue (f = 6), signaling a strong field-wide emphasis on cognitive processes. This is complemented by domain-specific contributions in journals such as the International Journal for Technology in Mathematics Education (f = 4) and Red-Revista De Educacion A Distancia (f = 3). Furthermore, a cluster of journals ranging from the Journal of Computer Assisted Learning to Frontiers in Psychology each published smaller sets of studies (f = 2), illustrating that the intersection of argumentative thinking, technology, and mathematics is gaining traction across diverse fields including psychology, linguistics, and higher education.

Distributions of Research Methods and Designs in the Studies

Based on the data collected, the distribution of studies conducted on argumentation thinking skills "critical thinking skills-ai-maths-tech" according to research methods and patterns is shown in Table 4.

Table 4. Research method and design distributions in the studies

Research Methods	f	Research Design	f	Code of the Study
Quantitative	11	Experimental	11	ACT3, ACT7, ACT1, ACT8, ACT12,ACT13,ACT14, AMATH1, AMATH2, ATEC1, ATEC2
Qualitative	2	Action Research	1	ACT11
		Ethnography	1	ACT5
Mixed	11	Explanatory sequential	7	ACT2, ACT4, ACT10, ACT9, ACT18, ACT19, ACT20
		Exploratory sequential	4	ACT6, ACT15, ACT16, ACT17
Total	24		24	

The methodological landscape of studies within the ACT, ATECH, and AMATH themes reveals a decisive preference for empirical validation over purely descriptive exploration. As detailed in Table 5, the field is dominated by an equal prominence of mixed-method and quantitative designs (f = 11 each), which collectively constitute the vast majority of the research. Notably, the quantitative studies are exclusively experimental in nature, while mixed-method approaches heavily prioritize explanatory sequential designs (f = 7). This pattern suggests that researchers are primarily focused on measuring the causal impact of specific interventions such as AI tools or mathematical reasoning models on students' skills, rather than merely observing the process. In stark contrast, purely qualitative inquiries, such as action research or ethnography, remain significantly underrepresented (f = 2), highlighting a potential gap in in-depth, naturalistic investigations within these domains.

Distributions of Sampling Methods, Levels and size in the Studies

Based on the data collected, the sampling method, level and size distributions of the studies are shown in Tables 5- 6 and 7.

Table 5. Distributions of Sampling Methods, Levels and Size

Sampling Methods	f	Types of Sampling	f	Coding of Study
Non-Probability sampling methods	23	Convenience / purposive	23	ACT3, ACT7, ACT1, ACT8, ACT2, ACT4, ACT10, ACT9, ACT11, ACT12, ACT13, ACT124, ACT15, ACT16, ACT17, ACT18, ACT19, ACT20, ACT5, AMATH1, AMATH2, ATEC1, ATEC2
Probability sampling		Simple random sampling	1	ACT6
Total	24		24	

As presented in Table 5, the vast majority of the studies utilized non-probability sampling methods (f = 23), specifically convenience and purposive sampling. This dominance suggests that researchers prioritized accessibility and practicality when selecting participants. Given the logistical challenges and strict permissions required for probability sampling in educational settings, non-probability methods appear to be a more feasible option for these studies. In contrast, simple random sampling was observed in only one study (f = 1), highlighting the rarity of purely randomized designs in this specific research context.

Table 6. Distributions sampling levels in the studies

Levels	f	Study Code
Secondary Students	3	ACT2, ACT6, ACT20
High Students	9	ACT11, ACT10, ACT9, ACT17, ACT18, ACT17, ACT1, AMATH2, ATECH2
Primary Students	2	ACT8, ACT5
Universty Students	10	ACT3, ACT7, ACT4, ACT1, ACT12, ACT13, ACT14, ACT15, ACT16, ATECH1
Total	24	

The distribution of study samples within the ATECH, AMATH, and ACT themes reveals a distinct upward skew towards higher educational levels. As evident from the data, the research landscape is heavily dominated by older student groups, specifically university (f = 10) and high school students (f = 9). In significant contrast, studies involving younger learners middle school (f = 3) and primary school students (f = 2) remain sparse. This disparity suggests that current literature tends to conceptualize argumentative thinking primarily as an advanced cognitive skill suitable for older adolescents and adults, thereby leaving a notable gap in understanding how these critical skills can be cultivated during early childhood and basic education stages.

Table 7. Distributions Sample Size in the Studies

Sample Size	f	Study Code
1-10	2	ACT1 ACT5
11-30	7	ACT3, ACT7, ACT4, ACT1, ACT12, ACT13, ACT14
31-100	15	ACT2, ACT6, ACT8, ACT10, ACT9, ACT11, ACT15, ACT16, ACT17, ACT19, ACT20, AMATH1, AMATH2, ATEC1, ATEC2
101-300	-	
301- 1000	-	
More than 1000	-	
Total	24	

The analysis of sample sizes within the ACT, AMATH, and ATECH themes reveals a clear operational preference for small-to-medium scale cohorts, likely reflecting the classroom-based nature of these interventions. As shown in Table 8, the vast majority of studies cluster within the 31–100 participant range (f = 15). This specific bracket typically corresponds to single or dual classroom sizes, which allows researchers to manage experimental conditions effectively while maintaining sufficient statistical power. Smaller sample groups, such as 11–30 (f = 7) and 1–10 (f = 2), are less frequent, suggesting that while pilot or case studies exist, the primary research thrust is directed towards verifying pedagogical outcomes in standard educational settings

Distributions of Data Collection Tools in the Studies

Based on the data collected, the distributions of the data collection tools of the studies conducted on argumentation thinking skills "critical thinking skills" are shown in Table 8.

Table 8. Distributions of Data Collection Tools in the Studies

Data Collection Tools	f	Coding
Scale	13	ACT12, ACT13, ACT14, ACT15, ACT16, ACT17, ACT18, ACT19, ACT20, AMATH1, AMATH2, ATEC1, ATEC2
Survey likert	5	ACT3, ACT7, ACT4, ACT1, ACT10
Semi- structured	7	ACT7, ACT1, ACT8, ACT2, ACT4, ACT11, ACT9
Document	3	ACT5, ACT11, ACT6
Total	28	

The profile of data collection instrumentation within the ACT, AMATH, and ATECH themes reflects a strong orientation towards standardized quantification. As detailed in Table 9, the field is heavily reliant on structured measurement tools, specifically scales (f = 13) and Likert-type surveys (f = 5). This dominance aligns with the prevalence of experimental designs noted earlier, as researchers prioritize tools that yield measurable, statistically analyzable data to validate skill acquisition. While quantitative metrics prevail, qualitative insights are primarily gathered through structured interviews (f = 7), further reinforcing a methodological preference for controlled and systematic data gathering over less structured approaches like document review (f = 3).

Distribution of Data Analysis in the Studies

Based on the data collection analysis distributions of the studies conducted on argumentation thinking skills "critical thinking skills" are shown in Tables 9-10 and 11.

Table 9. Distributions of qualitative analysis in the studies

Qualitative Data Analysis	f	Study Code
Content Analysis	5	ACT1, ACT5, ACT2, ACT3, ACT4
Descriptive Analysis	15	ACT6, ACT7, ACT8, ACT9, ACT10,ACT11, ACT12,ACT13, ACT14,ACT15,ACT16, ACT17,ACT18,ACT19,ACT20
Total	20	

The approach to qualitative data analysis within the ACT, AMATH, and ATECH themes reflects a methodological inclination towards summarizing broad patterns rather than deep textual deconstruction. As shown in Table 9, descriptive analysis (f = 15) emerges as the predominant technique, significantly outpacing content analysis (f = 5). This preference suggests that researchers are primarily concerned with organizing and presenting the data to depict the general characteristics of the findings, likely to complement the quantitative measures mentioned earlier. The relatively lower use of content analysis indicates that studies focusing on the latent meanings or complex conceptual structures within participants' responses are less common in the current literature.

Table 10. Distributions of descriptive statisticsl analysis in studies

Descriptive Statistics	f	Study Code
Frequency and Percentage	15	ACT6, ACT7, ACT8, ACT9, ACT10, ACT11, ACT12, ACT13,ACT14, ACT15, ACT16, ACT17, ACT18, ACT19, ACT20
Arithmetic mean, standard deviation	9	ACT12, ACT13, ACT14, ACT15, ACT16, AMATH1, AMATH2, ATECH1, ATECH2
Total	24	

The examination of descriptive statistical reporting within the ACT, AMATH, and ATECH themes indicates a dual focus on categorical profiling and performance measurement. As shown in Table 11, the most prevalent method is frequency and percentage calculations (f = 15). This widespread use suggests that researchers prioritize establishing clear demographic or categorical baselines for their samples. Following this, the use of the arithmetic mean (f = 9) reflects the necessity of analyzing continuous data derived from the scales and experimental designs mentioned earlier, allowing for the assessment of central tendencies in student performance.

Table 11. Distributions of Inferential Statistics in Studies

Inferential Statistics	f	Study Code
t-test	9	ACT12,ACT13,ACT14,ACT15,ACT16,ACT7,ACT8, ACT9,ACT10
Anova/ Ancova	13	ACT12, ACT13, ACT14, ACT15, ACT16, ACT3, ACT4, ACT5, ACT6, AMATH1, AMATH2, ATEC1, ATEC2
Correlation	6	ACT6, ACT6, ACT7, ACT8, ACT9, ACT10
Non Parametric tests	6	ACT6, ACT6, ACT7, ACT8, ACT9, ACT10
Regression	2	ACT9,ACT20
Total	36	

The landscape of inferential statistics within the ACT, AMATH, and ATECH themes strongly corroborates the experimental focus of the field. As outlined in Table 11, there is a clear methodological preference for group comparison analyses over predictive modeling. ANOVA/ANCOVA (f = 13) and t-tests (f = 9) emerge as the most frequently employed techniques. This dominance is consistent with the widespread use of quasi-experimental designs noted earlier, where researchers aim to determine the statistical significance of interventions on skill development. The notably high usage of ANOVA/ANCOVA suggests a commitment to methodological rigor, indicating that researchers frequently control for covariates (such as pre-test scores) to isolate treatment effects. In contrast, predictive analyses like regression (f = 2) are minimal, further confirming that the primary research objective in this domain is establishing causality rather than forecasting outcomes.

Findings Regarding the Theme of Argumentation Thinking Skills “Writing and Language Skills”

Distribution of studies by year

Based on the data collected, the distribution of studies on argumentation thinking skills, "writing and language skills (AWL) " by year is shown in Table 12.

Table 12. Distribution of studies by year

Years	Study Code	f
2014	AWL23, AWL24, AWL25	3
2015	AWL26,AWL27, AWL38	3
2016	AWL28,AWL29,AWL30	3
2017	AWL31, AWL32, AWL37	3
2018	AWL33, AWL34, AWL30	3
2019	AWL 35, AWL36	2
2020	AWL37, AWL38, AWL40,AWL43	4
2021	AWL3,AW4,AWL5,AWL6,AWL7,AWL8,AWL9,AWL42	8
2022	AWL10,AWL11,AWL12,AWL13	4
2023	AWL14,AWL15,AWL16,AWL17,AWL18,AWL19,AWL20, AWL21,AWL41	9
2024	AWL22, AWL39	2
	Total	43

The temporal evolution of research on Argumentative Writing and Language (AWL), as detailed in Table 12, illustrates a distinct dichotomy between the first and second halves of the examined decade. The early period (2014–2019) was characterized by a stagnant but stable interest, with publication outputs consistently hovering around low frequencies (f = 2–3 annually). However, a significant shift is observed starting in 2020, marking the onset of a surge in academic production. This upward trajectory culminated in notable peaks during 2021 (f = 8) and 2023 (f = 9). This rapid proliferation in recent years suggests a renewed scholarly emphasis on written argumentation, likely driven by the increasing necessity of high-level literacy and text-based communication skills in contemporary educational contexts

Table 13. Distribution of journals in studies

Journals	Study Code	f
Readıng And Writing	AWL1, AWL3, AWL34, AWL38	4
Innovatıon In Language Learnıng and Teachıng	AWL2, AWL5, AWL6,AWL7	4
Educatıo Sıglo Xxı	AWL8,AWL9,AWL10	3
Gıst-Educatıon and Learnıng Research	AWL11, AWL12, AWL34,AWL43	4
Estudıos Fılologıcos	AWL13, AWL14, AWL35	3
Englısh Teachıng-Practıce and critigu	AWL15, AWL16, AWL39,AWL40	4
Assessıng Wrıtıng	AWL17,AWL18,AWL19,AWL42	4
Asian-Pacific Journal Of Second and Foreıgn Language Educatıon	AWL20, AWL21, AWL22, AWL24, AWL26, AWL27, AWL28	7
Red-Revısta De Educacıon a Dıstancıa	AWL25, AWL26, AWL27, AWL28, AWL29, AWL30, AWL31,AWL41	8
Frontıers in Psychology	AWL36,AWL37	2
Toplam		43

The dissemination landscape of research on Argumentative Writing and Language (AWL) is characterized by a multidisciplinary breadth rather than being confined to a single specialized niche. As detailed in Table 13, the leading publication outlets reflect diverse academic focuses. RED–Revista de Educación a Distancia (f = 8) and the Asian-Pacific Journal of Second and Foreign Language Education (f = 7) emerge as the most prominent platforms, highlighting the strong intersection of this theme with distance education and second language acquisition. Furthermore, a significant cluster of studies is published in journals specifically dedicated to literacy and assessment, such as Assessing Writing and Reading and Writing (f = 4 each). The presence of journals like Frontiers in Psychology (f = 2) further underscores the cross-disciplinary relevance of the topic, bridging educational practice with psychological inquiry.

Distributions of Research Method and Design in the Studies

Based on the data collected, Table 14 presents the distribution of studies on AWL according to their research methods and designs.

Table 14. Distributions of research method and design in studies

Research Method	f	Research Design	f	Study Code
Quantitative	21	Experimental	16	AWL13, AWL5, AWL7, AWL9, AWL6, AWL17, AWL22, AWL13, AWL18, AWL42, AWL23,AWL24,AWL25, AWL31,AWL32,AWL33
		Relational	4	AWL14, AWL2,AWL30,AWL43
		longitudinal	1	AWL8
Mixed	17	Embedded	12	AWL12, AWL4, AWL10,AWL5,AWL16, AWL26,AWL27,AWL28,AWL29,AWL34,AWL35, AWL41
		Convergent	5	AWL17,AWL20, AWL21, AWL22, AWL23
Qualitative	5	Case Study	5	AWL36, AWL37, AWL38,AWL39,AWL40
Total	43		43	

The methodological architecture of research on Argumentative Writing and Language (AWL) demonstrates a clear orientation towards empirical intervention and impact assessment. As outlined in Table 14, the field is dominated by quantitative approaches (f = 21), heavily anchored by experimental designs (f = 16). This prevalence indicates that the primary scholarly objective is to rigorously test the efficacy of specific instructional strategies or writing interventions. Complementing this, mixed-method designs (f = 17)—particularly embedded designs (f = 12)—play a substantial role. This suggests that researchers frequently seek to bolster their statistical findings with qualitative data to better understand the processes underlying writing development. In contrast, purely qualitative case studies (f = 5) and longitudinal designs (f = 1) remain scarce, highlighting a significant gap in research regarding the long-term developmental trajectories of argumentative writing skills.

Distributions of Sampling Method Type, Level and size in the Studies

Based on the data collectedt, the sampling method type, level and size distributions of the studies related to AWL are shown in Tables 15-16-17.

Table 15. Distributions of Sampling Method Type in the Studies

Sampling Method	f	Sampling Type	f	Study Code
Non-Probability sampling methods	31	Convenience Sampling	31	AWL13, AWL5, AWL7, AWL9, AWL6, AWL17, AWL22, AWL13, AWL18, AWL12,AWL4, AWL10, AWL5, AWL16, AWL19, AWL23, AWL24, AWL25, AWL26, AWL27, AWL28, AWL29, AWL30, AWL31, AWL32, AWL33, AWL34, AWL35, AWL41, AWL42, AWL43
	12	Purposive sampling	12	AWL14, AWL8, AWL2, AWL17, AWL20, AWL21, AWL22, AWEL36, AWL37,AWL38,AWL39,AWL40
Toplam	43		43	

As presented in Table 15, the sampling strategies employed within the AWL theme were exclusively non-probability methods. The vast majority of the studies (f = 31) utilized convenience or mixed purposive approaches, while the remaining studies (f = 12) relied specifically on purposive sampling. This distribution highlights a methodological preference for accessibility and feasibility in educational contexts. Given the logistical challenges of implementing randomized selection in school settings, researchers appear to prioritize accessible participant groups (convenience) or specific target populations (purposive) over probability sampling.

Table 16. Distributions of sample level in the studies

Sample Level	f	Study Code
Middle school students	5	AWL22, AWL13, AWL18, AWL12, AWL4
High school students	12	AWL14,AWL8, AWL2, AWL17,AWL20, AWL21, AWL22, AWEL36,AWL37,AWL38,AWL39,AWL40
Primary school students	7	AWEL36,AWL37,AWL26,AWL27,AWL28,AWL29,AWL30
University students	19	AWL41,AWL31,AWL32,AWL33,AWL34,AWL35, AWL23,AWL24,AWL25, AWL13, AWL5, AWL7, AWL9, AWL6, AWL17, AWL22, AWL41,AWL42,AWL43,AWL13,
Total	43	

The sampling strategies employed within the AWL theme reflect a strong methodological commitment to representativeness, while also accommodating the need for targeted participant selection. As shown in Table 15, simple random sampling (f = 31) is the dominant approach, constituting the majority of the research. This prevalence suggests that researchers are acutely concerned with minimizing selection bias to ensure that findings on writing skills are generalizable to broader student populations. However, unlike other domains, purposive sampling (f = 12) also maintains a significant presence. This indicates that a subset of studies deliberately targets specific groups—likely based on writing proficiency levels or linguistic backgrounds—to align with the qualitative or mixed-method inquiries identified earlier.

Table 17. Distributions of sample size in the studies

Sample Size	f	Study Code
1-10	2	AWL9, AWL6
11-30	5	AWL26, AWL27,AWL28,AWL29,AWL30
31-100	26	AWL31,AWL32,AWL33,AWL34,AWL35,AWL23,AWL24,AWL25, AWL12, AWL16,AWL13, AWL22, AWL13, AWL14,AWL8, AWL2, AWL17,AWL20, AWL21, AWL22, AWL36,AWL37,AWL38,AWL41,AWL42,AWL43
101-300	10	AWL1,AWL2,AWL3,AWL4,AWL5,AWL39,AWL40,AWL10,AWL11,AWL15
301- 1000	-	
1000'den fazla	-	
Total	43	

The distribution of sample sizes within the AWL theme reveals a distinct operational preference for classroom-scale cohorts. As presented in Table 17, the most frequently selected sample size falls within the 31–100 range (f = 26). This specific bracket typically corresponds to standard class sizes (often involving one experimental and one control group), which directly supports the experimental and quasi-experimental designs previously identified as dominant in this field. Larger sample sizes of 101–300 (f =

10) are also notable, indicating a subset of research aiming for higher statistical power and broader generalizability. Conversely, small-group or individual case studies (1–30 participants) are significantly less common, reinforcing the trend that the field prioritizes collective pedagogical outcomes over idiosyncratic individual trajectories.

Distributions of Data Collection Tools in the Studies

Based on the data collected, The disributions of the data collection tools of the studies on AWL is shown in Table 18.

Table 18. Data collection tools in the studies

Data Collection Tools	f	Study Code
Scale	36	AWL31,AWL32,AWL33,AWL34,AWL35,AWL23, AWL24,AWL25,AWL13,AWL5,AWL7,AWL9,AWL6, AWL17,AWL22,AWL13,AWL14,AWL8,AWL2,AWL17, AWL20,AWL21,AWL22,AWL1,AWL2,AWL3,AWL4, AWL5,AWL39,AWL40,AWL10,AWL11,AWL15, AWL41,AWL42,AWL43
Survey Likert	12	AWL17,AWL20,AWL21,AWL22, AWEL36,AWL37,AWL38,AWL1, AWL2,AWL3,AWL4,AWL5
Semi-Structured interview	17	AWL22,AWL13,AWL14,AWL8,AWL2, AWL17,AWL20,AWL21,AWL22,AWEL36, AWL37,AWL38, AWL1,AWL2,AWL3,AWL4,AWL5
Total	65	

The profile of data collection instrumentation within the AWL theme reflects a methodological strategy that prioritizes triangulation—combining quantitative measurement with qualitative depth. As detailed in Table 18, the field is heavily anchored by standardized tools, specifically scales (f = 36) and Likert-type surveys (f = 12). This predominance aligns with the strong experimental focus noted earlier, highlighting a need to objectively quantify writing proficiency and argumentative skills. However, the significant integration of semi-structured interviews (f = 17) reveals a critical complementary trend: researchers are actively seeking to corroborate statistical outcomes with participants' subjective insights and experiences, thereby enriching the interpretative validity of the findings.

Distributions of Data Analysis in Studies

Based on the data collected, the findings regarding the data analysis distributions of the studies on AWL shown in Tables 19-20-21***.***

Table 19. Distributions of qualitative data analysis in the studies

Qualitative Data Analysis	f	Study Code
Content Analysis	5	AWL36, AWL37, AWL38,AWL39,AWL40
Descriptive Analysis	20	AWL1,AWL31,AWL32,AWL33,AWL34,AWL35, AWL23,AWL24,AWL25, AWL13, AWL5, AWL7,AWL9,AWL6,AWL17, AWL22, AWL13,AWL41,AWL42,AWL43
Total	25	

The strategy for analyzing qualitative data within the AWL theme reveals a distinct methodological preference for broad thematic summarization over granular text deconstruction. As presented in Table 19, descriptive analysis (f = 20) is the overwhelmingly dominant technique, surpassing content analysis (f = 5) by a significant margin. This distribution suggests that researchers primarily utilize qualitative data to depict general patterns or to categorize participant responses in a way that complements quantitative findings, rather than engaging in deep, latent coding of the textual material. The scarcity of content analysis implies that the field currently places less emphasis on uncovering complex, hidden meanings within student arguments, focusing instead on observable and explicitly stated characteristics.

Table 20. Distributions of descriptive statistics data analysis in the studies

Descriptive Statistics	f	Study Code
Arithmetic mean, standard deviation	23	AWL4,AWL10,AWL11,AWL12,AWL22, AWL13, AWL14,AWL8, AWL2, AWL16,AWL6,AWL17, AWL20, AWL18,AWL19,AWL21, AWL22, AWL36,AWL37,AWL38,AWL41,AWL42,AWL43
Frequency and Percentage	20	AWL1,AWL2,AWL3,AWL31,AWL32,AWL33,AWL34,AWL35, AWL23,AWL24,AWL25, AWL13, AWL5, AWL7, AWL9, AWL6, AWL17,AWL22, AWL15,AWL42,
Total	43	

The selection of descriptive statistical methods within the AWL theme underscores the field's preoccupation with measuring performance levels. As shown in Table 20, the arithmetic mean (f = 23) emerges as the most frequently applied metric, slightly surpassing frequency and percentage calculations (f = 20). This hierarchy aligns perfectly with the extensive use of scales and experimental designs identified in previous sections. Since the primary goal of these studies is often to quantify writing proficiency or argumentative quality, researchers rely heavily on mean scores to determine central tendencies in student performance. Frequency data, while also prominent, serves the complementary role of characterizing the demographic profiles of these study groups.

Table 21. Distributions of ınferential statistics data analysis in the studies

Inferentia Statistics	f	Study Code
t test	18	AWL13, AWL5, AWL7, AWL9, AWL6, AWL17, AWL22, AWL13, AWL18, AWL23,AWL24,AWL25, AWL31,AWL32,AWL33,AWL41,AWL42,AWL43
Anova/ Ancova	11	AWL12, AWL4, AWL10,AWL5,AWL16, AWL26,AWL27,AWL28,AWL29,AWL34,AWL35
Correlation	4	AWL14,AWL8, AWL2,AWL30
Non Parametric tests	5	AWL13, AWL18, AWL23,AWL24,AWL25
Regression	4	AWL17,AWL20, AWL21, AWL22
Mancova	2	AWL34,AWL35
Total	44	

The profile of inferential statistics within the AWL theme provides further evidence of the field's reliance on comparative experimental frameworks. As detailed in Table 21, independent and paired samples t-tests (f = 18) are the most widely utilized method, followed significantly by ANOVA/ANCOVA (f = 11). This heavy concentration on difference-testing statistics directly mirrors the prevalence of quasi-experimental designs identified earlier, where the primary objective is to assess the magnitude of change between control and experimental groups. In contrast, more complex relational or predictive analyses—such as correlation (f = 4), regression (f = 4), and multivariate methods like MANCOVA (f = 2)—are utilized far less frequently. This suggests that current research prioritizes establishing the immediate effectiveness of interventions over exploring intricate multivariate relationships or predictive models.

DISCUSSION AND CONCLUSION

This study systematically analyzed trends in argumentative thinking skills across 2014–2024, organizing the synthesis into two interrelated themes: first; ACT–ATECH–AMATH (critical thinking–AI/technology–mathematics) and second; AWL (argumentative writing and language). Treating these as complementary lenses allows the field's conceptual and methodological patterns to be read as a coherent whole rather than parallel strands.

Within the ACT–ATECH–AMATH theme, publications concentrated in 2021, with an ongoing upward tendency in 2024, a trajectory consistent with the post-pandemic emphasis on digital literacy and blended/online learning (Redecker, 2017). Prior work positions argumentation as a vehicle for critical thinking and increasingly integrated into STEM contexts (Osborne et. al, 2004; Erduran & Jiménez-Aleixandre, 2012). Methodologically, quantitative and mixed-methods designs often experimental predominated, reflecting a focus on intervention effects alongside process-level insights. Samples were mainly high-school and university populations, indicating a need to extend rigorous designs to earlier grades where foundational argumentation can be scaffolded developmentally (Kuhn & Udell, 2003). Inferential analyses were largely group-comparison oriented with comparatively limited use of multivariate modeling, which constrains the field's capacity to capture the complex, interacting determinants of argumentation.

Turning to the AWL theme, publications peaked in 2023 and 2021. A substantial share targeted reading–writing competencies, aligning with the view that argumentation is both a higher-order thinking skill and a core component of language development (Kuhn & Udell, 2003; Andrews, 2010). Integrating argumentative practices into literacy tasks fosters evidence evaluation, coherent argument construction, and persuasive expression, strengthening overall communicative competence. The strong link between argumentation and critical thinking was again evident (Nussbaum & Sinatra, 2003; Rapanta et al., 2013). In terms of analytic practice, AWL studies were dominated by t-tests and ANOVA/ANCOVA, followed by more limited use of non-parametric tests, correlation, regression, and MANCOVA, reinforcing the broader pattern of emphasizing group contrasts over multivariate explanations.

A cross-cutting pattern across both themes is the growing interest in connecting argumentation with artificial intelligence (AI) and mathematics. Educational AI is increasingly framed as a promising means to support data-driven problem-solving and evidence-based reasoning across STEM disciplines, thereby intensifying opportunities to elicit and assess argumentation in authentic tasks (Sain, 2024; Jiménez-Aleixandre & Erduran, 2008). Looking forward, the intersection of AI and technology (ATECH) with critical thinking (ACT) necessitates a paradigm shift from viewing technology as a mere delivery tool to recognizing it as an active 'argumentation partner.' The rapid integration of Generative AI (e.g., ChatGPT) into educational settings poses new challenges and opportunities. For instance, future pedagogical models might employ AI agents not just to grade arguments, but to act as 'Devil's Advocates' that expose students to counter-arguments in real-time, thereby forcing deeper cognitive processing. Consequently, the definition of 'argumentative thinking' must evolve to include 'algorithmic literacy'—the ability to critically evaluate machine-generated arguments for bias and hallucination. The temporal clustering of publications in 2021 and 2023 likely reflects the combined influence of pandemic-accelerated digitization and the contemporaneous prioritization of higher-order thinking in curricula and assessment (Hasnunidah et.al; 2020; Marni & Harsiatii 2019; Redecker, 2017).

Demographically and methodologically, the synthesis shows a clear upper-secondary/tertiary emphasis and a dominance of quantitative/mixed designs. Furthermore, the sampling strategies across both themes were overwhelmingly characterized by non-probability methods, specifically convenience and purposive sampling. Purely randomized probability sampling was exceptionally rare, observed in only a single study within the ACT theme. While this profile has yielded generalizable and reasonably dee insights, two gaps remain salient; the scarcity of longitudinal research limits understanding of how argumentative skills evolve over time; and the underuse of multivariate and discourse-analytic approaches restricts visibility into the mechanisms—cognitive, social, and contextual—through which argumentation develops. However, this scarcity should not be viewed merely as a methodological oversight, but rather as a reflection of the structural constraints inherent in educational research. Implementing longitudinal studies to track the developmental trajectory of argumentative thinking skills presents significant logistical challenges, primarily due to high attrition rates and curricular fragmentation across academic years. Similarly, the dominance of univariate statistics over multivariate models is often dictated by the practical reality of classroom-based research, where sample sizes are typically limited to intact classes, thus constraining the statistical power required for complex modeling.Notably, studies leveraging content analysis and descriptive statistics have provided valuable overviews, but broader use of discourse analysis and narrative inquiry would likely reveal additional layers in how claims, evidence, warrants, and counter-arguments are coordinated in real learning settings (Bekiroğlu & Ütkür-Güllühan, 2022; Erduran et al., 2015; Samson & Blanchard, 2012).

In conclusion, between 2014 and 2024, publications peaked in 2021 and 2023. Most studies used quantitative or mixed (often experimental) designs with high-school/university samples and relied on group-comparison statistics. These patterns align with the post-pandemic emphasis on digital literacy and higher-order thinking and support the view that argumentation plays a central role in critical thinking and language development. Future work would benefit from longitudinal designs, greater use of multivariate models, discourse-oriented qualitative analyses, and systematic AI–STEM integrations.

The primary contribution of this systematic review lies in exposing the imbalance between the rapid technological integration in argumentation studies and the stagnation of methodological diversity. By mapping the field’s trajectory over the last decade, this study establishes that while the thematic focus has successfully pivoted towards digital and AI-enhanced environments, the methodological apparatus has remained largely confined to short-term, quasi-experimental designs.

Consequently, this study offers a dual roadmap for future research: 1. Thematically, researchers must pivot from assessing basic digital skills to exploring 'human-AI collaborative argumentation,' specifically how students critically navigate algorithmic information. 2. Methodologically, there is an urgent imperative to adopt longitudinal designs that can capture the developmental 'stickiness' of argumentative skills beyond the immediate post-test window.

Ultimately, this review underscores that for argumentative thinking education to remain relevant in an AI-driven era, research practices must evolve to be as complex and dynamic as the technologies they seek to evaluate.

Recommendations

A principal limitation of this study is its scope, which is restricted to 67 argumentation-focused studies indexed in the Web of Science (WoS) between 2014 and 2024. Based on the findings, future research should expand longitudinal designs to earlier grade levels (especially primary), strengthen the use of multivariate statistics and discourse-oriented qualitative analyses, examine argumentation within AI/STEM integrations and digital learning environments, include samples that reflect socioeconomic and cultural diversity, employ advanced data approaches such as learning analytics, eye-tracking, and AI-assisted text analysis, and systematically report implementation fidelity and sustained effects.

Funding Statement: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of Interest Statement: The author(s) declare that there is no potential conflict of interest with respect to the research, authorship, and/or publication of this article.

Ethics Approval and Informed Consent Statement: The study, by its nature, does not require ethics committee approval.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Author Contribution Statement: First author Derya Bekiroğlu: 60% Second author (Assoc. Prof. Dr. Nur Ütkür Güllühan): 40%,

REFERENCES

Akhdinirwanto, R. W., Agustini, R., & Jatmiko, B. (2020). Problem-based learning with argumentation as a hypothetical model to increase the critical thinking skills for junior high school students. *Jurnal Pendidikan IPA Indonesia, 9*(3), 340–350. https://doi.org/10.15294/jpii.v9i3.19282

Andrews, R. (2010). *Argumentation in higher education*. London: Routledge.

Andriessen, J., & Baker, M. (2006). *Arguing to learn* (pp. 443-460). Na.

Bekiroğlu, D., & Güllühan, N. Ü. (2022). Trends of thinking skills in social studies and life studies courses: A meta-synthesis study. *Thinking Skills and Creativity*, *45*, 101099. https://doi.org/10.1016/j.tsc.2022.101099

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, *84*(3), 287-312. https://doi.org/10.1002/(SICI)1098-237X(200005)84:3<287::AID-SCE1>3.0.CO;2-A

Erduran, S., & Jiménez-Aleixandre, M. P. (2012). Argumentation in science education research: Perspectives from Europe. In *Science Education research and Practice in Europe*, (pp.253-289) Brill.

Erduran, S., Ozdem, Y. & Park, JY. Research trends on argumentation in science education: a journal content analysis from 1998–2014. *IJ STEM Ed* **2**, 5 (2015). https://doi.org/10.1186/s40594-015-0020-1

Erduran, S., Simon, S., & Osborne, J. (2006). Argumentation in science education: Perspectives from classroom-based research. In S. Erduran & M. P. Jiménez-Aleixandre (Eds.), *Argumentation in science education: Perspectives from classroom-based research* (pp. 1–27). Springer. https://doi.org/10.1007/978-1-4020-6670-2_1

Garcia-Mila, M., Felton, M., Miralda-Banda, A., & Castells, N. (2023). Pre-service teachers' knowledge, beliefs and predispositions to teach argumentation in their disciplines. *Journal of Education for Teaching*, *49*(4), 648-666. https://doi.org/10.1080/02607476.2022.2150536

Giri, V., Paily, M.U. (2020). Effect of Scientific Argumentation on the Development of Critical Thinking. *Sci & Educ* 29, 673–690. https://doi.org/10.1007/s11191-020-00120-y

Grant, M. J., & Booth, A. (2009). A typology of reviews: an analysis of 14 review types and associated methodologies. *Health İnformation & Libraries Journal*, *26*(2), 91-108.

Hasnunidah, N., Susilo, H., Irawati, M., & Suwono, H. (2020). The contribution of argumentation and critical thinking skills on students' concept understanding in different learning models. *Journal of University Teaching & Learning Practice*, *17*(1), 6. https://doi.org/10.53761/1.17.1.6

Jiménez-Aleixandre, M. P., & Erduran, S. (2008). Argumentation in science education: An overview. In S. Erduran & M. P. Jiménez-Aleixandre (Eds.), *Argumentation in science education: Perspectives from classroom-based research* (pp. 3–27). Springer.

Kienpointner, M. (2017). Rhetoric and argumentation. In *The Routledge Handbook of Critical Discourse*

Studies (pp. 228-241). Routledge.

Kuhn, D. (1992). Thinking as argument. *Harvard Educational Review*, *62*(2), 155-179. https://doi.org/10.1080/23735082.2015.994254

Kuhn, D., & Udell, W. (2003). The development of argument skills. *Child Development*, *74*(5), 1245–1260. https://doi.org/10.1111/1467-8624.00605

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Marni, S., & Harsiati, T. (2019). Critical thinking patterns of first-year students in argumentative essay. *Journal for the Education of Gifted Young Scientists*, *7*(3), 683-697. https://doi.org/10.17478/jegys.605324

Noroozi, O., Banihashem, S. K., Biemans, H. J., Smits, M., Vervoort, M. T., & Verbaan, C. L. (2023). Design, implementation, and evaluation of an online supported peer feedback module to enhance students' argumentative essay quality. *Education and Information Technologies*, *28*(10), 12757-12784. https://doi.org/10.1007/s10639-023-11683-y

Nussbaum, E. M., & Sinatra, G. M. (2003). Argument and conceptual engagement. *Contemporary Educational Psychology*, *28*(3), 384–395. https://doi.org/10.1016/S0361-476X(02)00038-3

Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science*, *328*(5977), 463-466. DOI: 10.1126/science.1183944

Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching*, *41*(10), 994-1020. https://doi.org/10.1002/tea.20035

Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argument in school science. *Journal of Research in Science Teaching, 41*(10), 994–1020. https://doi.org/10.1002/tea.20035

Rapanta, C., García-Mila, M., & Gilabert, S. (2013). What is meant by argumentative competence? An integrative review of methods of analysis and assessment in education. *Review of Educational Research*, *83*(4), 483–520. https://doi.org/10.3102/0034654313487606

Redecker, C. (2017). *European framework for the digital competence of educators: DigCompEdu* (Y. Punie, Ed.) [JRC Science for Policy Report, EUR 28775 EN]. Publications Office of the European Union. https://doi.org/10.2760/15977

Sain, Z. H. (2024). Utılızıng artıfıcıal ıntellıgence ın educatıon to equıp students for the challenges of the 21st century . *New Era Internatıonal Journal Of Interdıscıplınary Socıal Researches*, *9*(23), 55–62. https://doi.org/10.5281/zenodo.11079969

Sampson, V., & Blanchard, M. R. (2012). Science teachers and scientific argumentation: Trends in views and practice. *Journal of Research in Science Teaching*, *49*(9), 1122-1148. https://doi.org/10.1016/j.tsc.2022.101099https://doi.org/10.1002/tea.21037

Samson, P. L., & Blanchard, M. R. (2012). Argumentation in science education: The role of teacher scaffolding. *Science Education, 96*(5), 935–955. https://doi.org/10.1002/sce.21000

Schwarz, B. B., & Baker, M. J. (2017). *Dialogue, argumentation and education: History, theory and practice*. Cambridge University Press.

Simon, S. (2008). Using Toulmin's argument pattern in the evaluation of argumentation in school science. *International Journal of Research & Method in Education*, *31*(3), 277-289. https://doi.org/10.1080/17437270802417176

Tindale, C. W. (2004). *Rhetorical argumentation: Principles of theory and practice*. Sage.

Tippett, C. (2009). Argumentation: The language of science. *Journal of Elementary Science Education*, *21*(1), 17-25. http://www.jstor.org/stable/43155841. Accessed 19 Oct. 2024.