The European Educational Researcher

http://www.eu-er.com

Improving Primary School Students' Socio-Scientific Argumentation Skills Through the Example Issue of Urban Heatwaves

Christoforaki Maria, Mavrikaki Evangelia & Galani Apostolia

National and Kapodistrian University of Athens, Greece

Abstract: The increasing frequency and intensity of urban heatwaves reshapes environmental and social realities globally. These phenomena not only challenge the resilience of urban ecosystems but also call attention to the importance of developing scientific literacy and reasoning skills at a primary level. Situated within these pressing socio-environmental conditions, this study examines the way in which primary school students engage in socio-scientific argumentation (SSA) when reasoning about urban heatwave mitigation. Drawing upon Toulmin's argumentation model as theoretical lens to explore both the structure and content of students' arguments and to trace how these evolve through targeted intervention, a qualitative-dominant design was employed involving 148 5th-grade students in Greece, with quantitative summaries used descriptively to illustrate trends and shifts in argumentation components. Data were collected through pre- and post-intervention digital questionnaires and structured worksheets. Students' written responses were analyzed using a qualitative content analysis, guided by a rubric that assessed both the structural coherence and content quality of their arguments, while frequencies and shift analysis provided descriptive insights into changes across demographic variables. The instructional intervention centered on the serious game Heatwave City, designed to immerse learners in authentic decision-making scenarios around urban sustainability. Findings revealed that pre-intervention arguments were largely fragmented and weakly supported by evidence, whereas post-intervention data indicated more consistent articulation of claims and increased use of relevant scientific information, particularly in questionnaire responses. Advances in reasoning and counterargumentation were more limited, with these components remaining largely at emergent levels. We conclude by offering recommendations for fostering students' engagement with SSA through sustained, dialogic, model-based, and game-enhanced learning environments.

Keywords: Primary education, scientific literacy, socio-scientific argumentation, urban heatwaves.
DOI: https://doi.org/10.31757/euer.912

Introduction

Rising global temperatures and the increasing frequency of extreme weather events continue to redefine environmental and social realities worldwide. Among these phenomena, urban heatwaves have emerged as one of the most critical challenges of contemporary urban life, disproportionately affecting vulnerable populations and exacerbating existing issues of energy consumption, health, and sustainability. In response to these escalating concerns, international policy frameworks, including the European Green Deal (European Commission, 2019) and the UN Sustainable Development Goals (United Nations, 2015), call for educational initiatives that cultivate scientifically literate citizens capable of making informed, evidence-based decisions about climate adaptation and mitigation. Within this context, science education is positioned not merely as a vehicle for conceptual understanding, but as a means for developing socio-scientific reasoning and agency in addressing real-world environmental problems.

Such goals align with the broader reform agenda for science education across Europe, which emphasizes inquiry-based learning, argumentation, and engagement with authentic socio-scientific issues (Osborne & Dillon, 2008; National Research Council, 2012). These pedagogical shifts reflect an understanding that the development of scientific literacy involves not only knowledge of scientific concepts, but also the ability to evaluate evidence, consider multiple perspectives, and construct reasoned arguments about complex societal issues. Central to this reform discourse is the notion of socio-scientific argumentation (SSA), which encompasses the cognitive and dialogic processes through which students generate, justify, and critique claims grounded in scientific evidence (Sadler & Zeidler, 2005; Erduran & Jiménez-Aleixandre, 2007).

However, despite the widespread endorsement of argumentation as a core practice in science education, empirical studies reveal that primary school students face significant challenges in constructing coherent, evidence-based arguments on socio-scientific issues (Khishfe, 2014; Osborne, 2010). Young learners often demonstrate fragmented reasoning, limited use of evidence, and difficulty integrating scientific principles into everyday contexts (Zeidler, 2014). These difficulties point to the need for carefully designed learning environments that scaffold argumentation processes and make space for dialogic interaction and reflective reasoning. Moreover, as recent studies have shown, students' engagement with SSA may vary depending on contextual and demographic factors such as gender and academic performance, suggesting that equitable instructional designs must account for such variations (Evagorou & Osborne, 2013).

Given these considerations, there arises a need to examine how primary students engage with socio-scientific argumentation in authentic environmental contexts and how targeted interventions might foster their reasoning competencies. Situated within this reform-oriented and theoretical landscape, the present study explores the development of 5th-grade students' SSA skills in the context of urban heatwave mitigation. Drawing upon Toulmin's Argumentation Model as an analytical framework, this research investigates not only students' initial SSA level, but also the extent to which these skills can be enhanced using a serious game designed to promote evidence-based reasoning and collaborative discourse. Specifically, the study addresses the following research questions:

(R.Q.1) Which SSA level (prior to intervention) are 5th-grade primary school students classified into regarding combating heatwaves in the urban environment?

(R.Q.2). To what extent can students' SSA level be improved in addressing urban heatwaves after intervention?

(R.Q.3). How do certain demographic factors (gender and school performance) affect 5th-grade primary school students' SSA level about counteracting urban heatwaves?

Theoretical framework

The increasing prevalence of global challenges has raised questions about how teachers can cultivate students' skills to engage with complex, ill-structured socio-scientific issues (SSIs) (Kolsto, 2006). Societies increasingly recognize that students represent the adults of tomorrow (Maloney, 2007) and that their ability to make well-informed decisions, grounded in reasoned argumentation, has far-reaching implications for individual and collective futures (Capkinoglu et al., 2020). Yet, despite this recognition, reasoning and deliberation in SSIs have only recently been acknowledged as explicit objectives within science education (Romine et al., 2017). In response, researchers have highlighted the importance of a specific form of argumentation, termed socio-scientific argumentation (SSA), which encompasses the construction, evaluation, and negotiation of claims within authentic, real-world contexts influenced by social norms, ethical values, and multifaceted constraints (Grooms et al., 2014; Kuhn, 2003).

Through engagement with SSA, learners are not only prompted to examine the merits and limitations of a proposition, but are also encouraged to explore alternative solutions, weigh competing perspectives, and grapple with the inherent complexities of socio-scientific issues (Zohar & Nemet, 2002; Dawson & Carson, 2020). Such engagement promotes deeper cognitive and conceptual competencies, while simultaneously addressing the broader demands of contemporary scientific literacy, which encompasses not only evolving scientific knowledge and technological advances, but also the ability to reason critically about issues situated at the intersection of science and society (Erduran et al., 2004).

Embedding principles of informal reasoning (Wu & Tsai, 2011), moral reasoning (Sadler, 2009), and socio-scientific reasoning (Kinslow et al., 2018) within SSA elucidates the cognitive mechanisms through which students construct, evaluate, and refine their arguments, while interrogating the interplay of evidence, values, and contextual factors. As such, SSA constitutes an integrative pedagogical scaffold, fostering both scientific literacy and the development of students as proactive, informed citizens.

Measuring the quality of SSA

The advancement of SSA skills requires systematic and theoretically informed approaches to evaluating argument quality. Among the key dimensions, the structural and content-related aspects of argumentation are widely recognized as critical components of reasoning competence. Frameworks such as Toulmin's Argumentation Pattern (TAP) provide a coherent analytical lens for identifying the elements essential to a well-structured argument: claim, grounds (or data), warrant, backing, qualifier, and rebuttal (Toulmin, 2000). In some interpretations, grounds and data are treated as distinct components, with grounds referring to the factual basis of the claim and data to the empirical or observational evidence supporting it (Erduran, Simon, & Osborne, 2004). These interdependent elements collectively determine the internal coherence, logical sufficiency, and overall persuasiveness of an argument.

The content dimension, in turn, evaluates the accuracy, depth, and conceptual validity of the scientific information embedded within the argument. High-quality arguments are characterized by scientifically accurate reasoning, the use of valid and contextually appropriate evidence, and the avoidance of misconceptions or oversimplifications that distort the complexity of socio-scientific issues (Sadler & Donnelly, 2006; Tal & Kedmi, 2006; Zohar & Nemet, 2002).

Christensen and Chang Rundgren's (2014) framework serves as the integrative basis for these criteria. Their model evaluates socio-scientific argumentation along two orthogonal axes: (1) structural complexity, reflecting the inclusion and interconnection of the Toulmin components, and (2) content quality, capturing the accuracy, validity, and justification of the scientific information. By combining these axes, the framework enables a multidimensional interpretation of students' reasoning, distinguishing between arguments that are structurally elaborate, but scientifically weak and those that are concise yet conceptually robust. This dual-perspective approach was used to guide both the design of the coding scheme and the interpretation of argument quality in this study.

Unlike conventional scientific argumentation, SSA requires attention to ethical, social, political, and economic dimensions, reflecting the multifaceted character of real-world decision-making. Thus, assessing SSA extends beyond factual correctness; it entails examining students' intellectual rigor, critical engagement with competing perspectives, and capacity to integrate diverse sources of evidence into coherent, context-sensitive arguments (Christensen & Chang Rundgren, 2014).

Students' difficulties in SSA

A central concern in contemporary science education is how primary students develop the skills to engage in socio-scientific argumentation (SSA), particularly when addressing complex, real-world issues such as urban heatwaves. Research consistently indicates that learners encounter significant challenges in constructing coherent and evidence-based arguments. These challenges manifest in multiple ways: (a) difficulty in selecting and integrating appropriate evidence, (b) limited ability to link evidence to claims through reasoning, and (c) struggles with engaging alternative perspectives via counterarguments (McNeill & Krajcik, 2012).

In the context of urban heatwaves, for instance, students may propose that planting a few trees in one neighborhood will eliminate the city's overall heat island effect, thereby overextending conclusions beyond the available evidence and neglecting the multifaceted nature of mitigation strategies (Vicedo-Cabrera et al., 2021; Zölch et al., 2019). Such reasoning patterns are often compounded by reliance on anecdotal observations or simplified conceptual models, which constrain the scientific rigor and credibility of students' arguments (Skoumios, 2023; Sadler, 2004; Zeidler, 2005).

Why do these difficulties persist at the primary level? One contributing factor is students' emerging disciplinary literacy. Limited mastery of scientific language and discourse conventions constrains their ability to construct logical connections and justify claims effectively (Moje et al., 2004). From a developmental perspective, primary students are typically transitioning through Piaget's concrete operational stage, where abstract reasoning is still developing, and rely heavily on socially mediated learning processes, as described by Vygotsky (1978). These cognitive and linguistic characteristics shape not only students' ability to formulate claims but also their engagement with evidence, consideration of alternative viewpoints, and integration of complex social and ethical dimensions.

The instructional implications of these challenges are significant. Effective development of SSA requires scaffolding that supports both cognitive and linguistic capacities, while embedding learning within meaningful, context-rich scenarios. Urban heatwaves provide a particularly suitable context for this work: as tangible manifestations of climate change, they are immediately relevant to students' everyday lives and involve interdisciplinary knowledge spanning meteorology, urban planning, public health, and social equity (Monroe et al., 2019; Sadler et al., 2017). Engaging with these scenarios allows students to articulate claims, justify them with evidence, consider alternative perspectives, and evaluate broader societal implications. Structured activities, such as guided debates, role-playing, and collaborative reflection, can promote these skills while simultaneously fostering scientific literacy and ethical reasoning.

Students' difficulties in SSA are not merely epistemic; they are intertwined with developmental, cognitive, and linguistic factors. The theoretical framework for this study, therefore, situates SSA within the interplay of reasoning, evidence, and context, highlighting the need for instructional designs that scaffold argument construction, support disciplinary literacy, and provide authentic, socially relevant problem spaces. Addressing these dimensions is essential to understand how 5th-grade students can develop robust socio-scientific argumentation skills in the context of urban heatwave mitigation, laying the foundation for informed, responsible, and proactive citizenship.

Urban heatwaves as a context for SSA

The challenges students face in socio-scientific argumentation are particularly pronounced when they engage with abstract or unfamiliar issues. As mentioned before, urban heatwaves provide a tangible, socially relevant, and scientifically rich context through which learners can bridge theoretical knowledge and real-world phenomena (Vicedo-Cabrera et al., 2021). Intensified by climate change and urbanization, these events have direct implications for public health, energy demand, and ecological sustainability, making them immediately relevant to students' daily lives and local environments.

Educational research underscores the pedagogical value of climate-related and environmental contexts for fostering reasoning about socio-scientific issues (Monroe et al., 2019; Stevenson et al., 2021; Shepardson et al., 2017). Urban heatwaves exemplify complex systems problems that integrate scientific, ethical, and socio-political dimensions. Learners engaging with such phenomena are prompted to construct arguments that combine meteorological knowledge, urban planning considerations, public health evidence, and principles of environmental justice. For instance, evaluating mitigation strategies, such as expanding green infrastructure, adopting reflective building materials, or redesigning urban layouts—requires students to weigh trade-offs, consider feasibility, and reflect on the equitable distribution of resources across socio-economic groups (Zölch et al., 2019).

These deliberations foreground the socio-political nature of science in society. Students are encouraged to examine whether proposed interventions disproportionately benefit affluent areas while neglecting marginalized communities, thereby cultivating reasoning that integrates equity, justice, and civic responsibility. Structured argumentation and reflective practices enable learners to develop scientific literacy alongside ethical awareness and democratic dispositions, equipping them to navigate complex socio-scientific issues thoughtfully (Zeidler, 2018). Urban heatwaves thus provide not only a scientifically engaging context but also a platform for integrating reasoning, values, and social responsibility, preparing primary-level students to become informed, critical, and empathetic citizens in a world increasingly shaped by climatic uncertainty.

Materials and methods

The study sample consisted of 148 fifth-grade students from three public primary schools in Greece (Table 1). A convenience sampling approach was employed, guided by the availability of schools willing to participate and practical considerations related to access and logistics. While convenience sampling inherently limits the generalizability of findings, it offers valuable opportunities for exploratory inquiry into the development of socio-scientific argumentation (SSA) within authentic educational contexts (Patton, 2015).

Gender and school performance were included as variables, given prior research suggesting their potential influence on students' argumentation skills and engagement with socio-scientific issues (Sadler, 2004; McNeill & Krajcik, 2012). School performance was assessed using a 3-point scale (1 = High, 2 = Medium, 3 = Low). To provide a conventional GPA-style metric, scores were inverted so that higher values indicate better performance (GPA = 4 − score). Male students (N = 80) had a mean school performance of 1.47 (SD = 0.50), corresponding to a mean GPA of 2.53. Female students (N = 68) had a mean score of 1.49 (SD = 0.59), corresponding to a mean GPA of 2.51. Overall, the total sample (N = 148) demonstrated a mean school performance of 1.48 (SD = 0.54) and a mean GPA of 2.52, indicating generally high academic achievement across the sample. Scores ranged from 1 to 2 for males and 1 to 3 for females, reflecting slightly greater variability among female students. All ethical standards were rigorously upheld. Informed parental consent was obtained for all participants, ensuring voluntary participation. The study received ethical approval from the Ethics and Deontology Committee for Research of [University] under Protocol Number 20566/5.3.2024 (Decision No. 150/21.3.2024).

Table 1
Description of the research sample.

Variable	Male (N=80)	Female (N=68)	Total (N=148)
Mean ± SD (Original Scale)	1.47 ± 0.50	1.49 ± 0.59	1.48 ± 0.54
Min – Max	1 – 2	1 – 3	1 – 3
Mean ± SD (GPA Scale)	2.53 ± 0.50	2.51 ± 0.59	2.52 ± 0.54
Gender Distribution	80 (54%)	68 (46%)	148 (100%)

Stages of the research

The study was conducted between February and June 2024 and was structured into three sequential phases, employing a mixed-methods qualitative design to assess and support students' socio-scientific argumentation (SSA) skills in the context of urban heatwave mitigation.

Phase 1: Pre-intervention assessment: The first phase aimed to establish students' baseline SSA competencies. Two complementary instruments were employed: (a) a digital questionnaire administered during a one-hour teaching session, and (b) structured worksheets completed over two teaching hours. These tools elicited students' existing understanding of urban heatwaves, their argumentative structure, and their use of evidence in reasoning about heatwave management strategies.

Phase 2: Intervention: The instructional intervention focused on the online serious game Heatwave City (https://platform.heatwaves-project.eu/resources/online_game/4), which simulates the complexities of managing urban heatwaves in urban environments. The game consists of four missions, each requiring students to make decisions that influence city temperature regulation, with the goal of maintaining a sustainable average of approximately 25°C. Each mission presents authentic socio-scientific scenarios, requiring students to weigh trade-offs among scientific evidence, ethical considerations, economic feasibility, and political implications.

Heatwave City has been validated in prior research as an effective tool for supporting students' decision-making and argumentation in complex socio-scientific contexts (Authors et al., 2024). Its design explicitly targets learning goals related to evidence-based reasoning, collaborative decision-making, and consideration of socio-ethical trade-offs, ensuring both content validity and instructional alignment.

Students were organized into small groups of four, each provided with a digital device (tablet or laptop) at individual workstations, fostering peer discussion and collaborative reasoning. The researcher acted as a facilitator, offering scientific clarification only when requested, while refraining from guiding game choices to preserve the autonomy of students' argumentation.

The intervention lasted approximately three teaching hours, allowing sufficient time for gameplay, intra-group discussion, and reflective analysis. After each mission, groups completed brief reflection logs documenting the rationale behind their decisions, disagreements, and alternative viewpoints. These logs were subsequently used to triangulate data with the post-intervention questionnaires and worksheets.

Phase 3: Post-intervention assessment: Three months after the intervention, the same instruments used in the pre-intervention phase were re-administered to evaluate changes in students' SSA performance. The follow-up questionnaire and worksheets were conducted with the same teaching time allocations to ensure consistency in data collection. This post-assessment focused on students' ability to construct claims, use evidence, reason effectively, and engage with counterarguments in the context of urban heatwave management. The study acknowledges that during this three-month interval, students may have experienced additional learning opportunities or classroom activities that could have influenced their argumentation skills. As such, this follow-up reflects a naturalistic assessment rather than a fully controlled longitudinal design. The post-assessment focused on students' ability to construct claims, use evidence, reason effectively, and engage with counterarguments in the context of urban heatwave management.

Overall, the intervention provided approximately six hours of structured and autonomous argumentation opportunities. Although brief, it allowed repeated engagement with socio-scientific decision-making, and reflection logs offered insights into students' reasoning processes. Limitations regarding external learning experiences and classroom influences are acknowledged and discussed.

Tools

The study followed a qualitative-dominant design. While students' written responses were systematically coded and analyzed qualitatively, descriptive statistics were employed to summarize frequencies and shifts in argument components, without making inferential comparisons or claiming experimental control. This approach enabled a comprehensive evaluation of SSA, combining depth of qualitative insights with rigor in quantitative assessment (Creswell & Plano Clark, 2018).

A pre- and post-intervention questionnaire administered via Google Forms served as the primary instrument. It included six items: two demographic questions (gender, school performance) and four open-ended items targeting students' SSA skills in relation to urban heatwaves (Table 2). Items were developed based on prior literature (Khishfe et al., 2023; Zeidler et al., 2005; Sadler, 2004), reviewed by three science education experts to ensure content validity, and piloted with 15 students to refine clarity and suitability.

The second research tool constituted structured worksheets, including two activities based on role-play and structured debate. These interactive tasks were designed to foster collaborative reasoning, perspective-taking, and justification of claims (Osborne, 2010; Zeidler & Sadler, 2008). Students worked in small groups of four, engaging in dialogue to simulate real-world decision-making processes. Detailed instructions guided group roles, discussion turns, and reflection logging to ensure consistency across classrooms.

Data analysis

The students' questionnaires and worksheets underwent pre-and post-content analysis (Berg, 2004), underpinned by two distinct rubrics, each incorporating graded evaluation criteria designed to assess (Table 3): (a) the structural coherence of students' written arguments, (b) the substantive content of these arguments.

Table 2

Description of open-ended items

Questions	Description and Intended Focus
Q1. What are the main challenges posed by heatwaves in urban environments, and how would you argue their significance compared to other urban issues?	*This question prompts to prioritize and justify why heatwaves are a significant challenge, encouraging them to compare and argue the relative importance of this issue.*
Q2. Why do you believe it is necessary (or not) to take action to modify urban environments in response to heatwaves? Present your arguments, including evidence or examples to support your position.	*This reformulation asks to present an argument either for or against acting, supporting their stance with evidence, to assess the strength of their reasoning.*
Q3. What specific solutions would you propose for reducing the impact of heatwaves through urban planning, and what arguments can you make to support their effectiveness?	*Students are encouraged not only to propose solutions, but also to back them up with arguments, including scientific evidence and case studies.*
Q4. In your opinion, what are the main weaknesses or challenges in the proposed methods to address heatwaves through urban development, and how would you argue for improving them?	*Students criticize existing approaches, supporting their critique with arguments and suggesting improvements.*

To evaluate the responses to the open-ended questions of the digital questionnaire and the worksheets, content and structure, were assessed separately using a predefined coding scheme, rather than through a composite score derived from the rubric shown in Table 3. Specifically, responses classified as Level 0 were assigned a score of 0, those at Level 1 received a score of 1, and those at Level 2 were given a score of 2.

Toulmin's full argumentation model (claim, data, warrant, backing, qualifier, rebuttal) provided the conceptual lens for this study, shaping how students' socio-scientific reasoning was theoretically understood. However, the analytical rubric operationalized four components, claim, evidence, reasoning, and counterargument, reflecting both the developmental characteristics of primary school students and the nature of the written data collected. Warrants and backings were analytically subsumed within the reasoning component, as students at this age level rarely articulate these elements explicitly and instead embed them implicitly within explanatory statements that connect evidence to claims (Erduran et al., 2004). Qualifiers were not coded as a separate category, given their limited presence in students' written responses and the focus of the study on core argumentative coherence rather than epistemic nuance.

Table 3
Graded evaluation criteria for content & structure of students' written arguments (Skoumios, 2023)

Structure Criteria		
Level 0	**Level 1**	**Level 2**
Does not propose a claim	Proposes an insufficient claim	Proposes a sufficient claim
Does not provide evidence	Provides insufficient evidence	Provides sufficient evidence
Does not provide reasoning	Proposes insufficient reasoning	Proposes sufficient reasoning
Does not provide a counterargument	Proposes an insufficient counterargument	Provides a sufficient counterargument
Content Criteria		
Level 0	**Level 1**	**Level 2**
Proposes an inappropriate claim	Proposes a partially appropriate claim	Proposes an appropriate claim
Provides inappropriate evidence	Provides partially appropriate evidence	Provides appropriate evidence
Proposes inappropriate reasoning	Proposes partially appropriate reasoning	Proposes appropriate reasoning
Proposals an inappropriate counterargument	Proposes a partially appropriate counterargument	Provides an appropriate counterargument

The coding process was conducted by an independent researcher trained on a preliminary set of approximately 20 worksheets to ensure familiarity with the rubric and consistency in application. To establish reliability, a subset of 30% of student responses was independently coded by a second trained researcher. Inter-rater agreement was evaluated using Cohen's kappa for each coding category across both dimensions (Table 4). In addition to reporting kappa values, the analysis included percentage agreement and a category-by-category examination of discrepancies, highlighting which rubric criteria were challenging to code consistently. Discrepancies were resolved through discussion, and consensus codes were applied to the full dataset.

Table 4
Inter-rater reliability for structural and content dimensions of students' responses.

Dimension	Category	Cohen's K	95% Confidence Interval	Percentage Agreement (%)
Structural	Claim	0.82	0.76 – 0.88	90
	Evidence	0.79	0.72 – 0.85	88
	Reasoning	0.81	0.75 – 0.87	89
	Counterargument	0.78	0.71 – 0.84	87
Content	Claim	0.80	0.74 – 0.86	89
	Evidence	0.77	0.70 – 0.83	86
	Reasoning	0.79	0.73 – 0.85	88
	Counterargument	0.75	0.68 – 0.82	85

The students' arguments met structure sufficiency if they encompassed the following elements: (a) a claim that directly addressed the initial question, (b) all the relevant evidence that could substantiate the claim, (c) a reasoning process that connected the evidence to the claim through the application of scientific theories and principles, (d) a rebuttal supported by adequate evidence and reasoning.

Evaluating the content, the appropriateness of each component was scrutinized, as well as the scientific validity within the context of school-based knowledge, regardless of the argument's overall sufficiency. The students' arguments were considered to fulfill the appropriateness criterion if they included: (a) a suitable claim, (b) appropriate supporting evidence, (c) reasoning that engaged relevant scientific principles to link the evidence to the claim, (d) a rebuttal with appropriate evidence and reasoning.

It is important to clarify that the analysis of students' arguments was conducted separately for content and structural dimensions, without integrating the two into a combined framework. This approach was adopted to capture the distinct and valuable contributions each component makes to understanding students' reasoning. Furthermore, the analysis was performed on an individual basis, enabling a detailed exploration of each student's developmental trajectory and revealing variations in the alignment between the quality of content and the structure of their arguments.

Below, there is an example (Table 5) of how a student response was analyzed in accordance with Table 3:

Worksheet activity: "As the urban planner of your city make an argument to convince the City Council about the best practice that should be followed to make the city resilient in heatwaves".

Student's response: "The best way to address heatwaves in our city is to create parks and plant trees. Trees perform photosynthesis, which means they will absorb more sunlight, thereby reducing the temperature".

Table 5
Student's argument evaluation

Structure	Content
A claim and supporting evidence are identified. The claim is sufficient, and the supporting evidence is adequate. However, there is no reasoning or counter-argument present.	Appropriate Claim and Suitable evidence are provided, but reasoning and counterargument are absent.
Claim: Level 2	**Claim:** Level 2
Evidence: Level 2	**Evidence:** Level 2
Reasoning: Level 0	**Reasoning:** Level 0
Counterargument: Level 0	**Counterargument:** Level 0

To further examine the developmental trajectory of students' SSA level, Shift Analysis was conducted, comparing pre- and post-intervention performance across the four argument components (claim, evidence, reasoning, counterargument) for both content and structure, separately for the questionnaire and the worksheets results.

To examine associations between students' demographic characteristics and their argumentation levels, cumulative link mixed models (CLMMs) were planned a priori. The dependent variable in these models was the students' argumentation level for each component, as an ordinal outcome, with higher values corresponding to more developed argumentation. Random intercepts for students and classrooms accounted for repeated measures and classroom-level clustering, while fixed effects included gender, school performance, and their interaction. Odds ratios derived from these models reflect the likelihood of a student achieving a higher cumulative argumentation category, providing a descriptive, inferential perspective on demographic trends that complements the primarily count- and shift-based analyses.

Results

Determining students' prior SSA level about mitigating urban heatwaves

Questionnaire

The pre-intervention analysis of students' written arguments revealed distinctive patterns in their content-level performance across the four components of argumentation: Claim, Evidence, Reasoning, and Counterargument, evaluated across four open-ended questions (Q1–Q4) (Table 6).

Across the claim and evidence components, students' responses were predominantly situated in the lower tiers of performance. In Q1, for instance, 75 students produced irrelevant or absent content (Level 0), 50 demonstrated partial adequacy (Level 1), and only 23 formulated fully developed arguments (Level 2). This distribution pattern remained largely stable through Q2 and Q3, but in Q4, a more cognitively demanding item, the number of Level 0 responses rose sharply to 100, while fully developed arguments decreased to 10.

The reasoning and counterargument components revealed an even more critical gap. Across all four questions, no students achieved Level 2, and the majority remained at Level 0 (ranging from approximately 100 in Q1 to 122 in Q4). A smaller subset achieved Level 1 (e.g., 48 in Q1 and Q3; 26 in Q4), indicating limited but emerging attempts to link claims with supporting evidence or to acknowledge opposing perspectives.

Patterns observed in the structural dimension (Table 6) broadly mirrored those in content but revealed important nuances in students' ability to organize and scaffold their ideas coherently. For the claim and evidence components, a modest level of structural awareness was evident. In Q1, approximately half of the students (n = 55 for claim; n = 60 for evidence) produced disorganized or absent structures (Level 0), while a comparable proportion achieved the highest level of structural coherence (Level 2; 58 and 55, respectively). This relative balance persisted through Q2 and Q3, but a decline was observed in Q4: Level 0 responses increased (n = 70 for Claim; n = 72 for Evidence), and fully coherent structures (Level 2) decreased to 38 and 36, respectively.

The reasoning and counterargument components, however, showed persistently low levels of structural sophistication. No student reached Level 2 in any question, and the vast majority remained at Level 0 (ranging from 100 to 124 per component). Only a minority achieved Level 1 (approximately 24–48 students per question), suggesting minimal structural attempts to connect reasoning with evidence or to integrate alternative viewpoints. In Q4, the most demanding task, Level 0 responses increased further (n = 124 for both components), underscoring the difficulty students faced in coordinating complex argumentative structures such as causal chains or rebuttal sequences.

Table 6

Assessment of the content and structure of students' responses to each questionnaire item prior to the intervention

Content

	Q1			Q2			Q3			Q4		
Components	**Pre_0**	**Pre_1**	**Pre_2**	**Pre_0**	**Pre_1**	**Pre_2**	**Pre_0**	**Pre_1**	**Pre_2**	**Pre_0**	**Pre_1**	**Pre_2**
Claim	75	50	23	80	32	36	90	35	23	100	38	10
Evidence	75	50	23	80	32	36	90	35	23	100	38	10
Reasoning	100	48	0	95	53	0	100	48	0	122	26	0
Counterargument	100	48	0	95	53	0	100	48	0	122	26	0

Structure

	Q1			Q2			Q3			Q4		
Components	**Pre_0**	**Pre_1**	**Pre_2**	**Pre_0**	**Pre_1**	**Pre_2**	**Pre_0**	**Pre_1**	**Pre_2**	**Pre_0**	**Pre_1**	**Pre_2**
Claim	55	35	58	50	38	60	70	26	52	70	40	38
Evidence	60	33	55	52	34	62	66	30	52	72	40	36
Reasoning	100	48	0	100	48	0	110	38	0	124	24	0
Counterargument	100	48	0	100	48	0	110	38	0	124	24	0

Low Medium High

Worksheets

The analysis of pre-intervention worksheets reveals differential performance across the content and structure components of student argumentation. Figure 1 presents the exact distribution of responses across rubric levels for each component, thereby ensuring clarity and consistency in reporting.

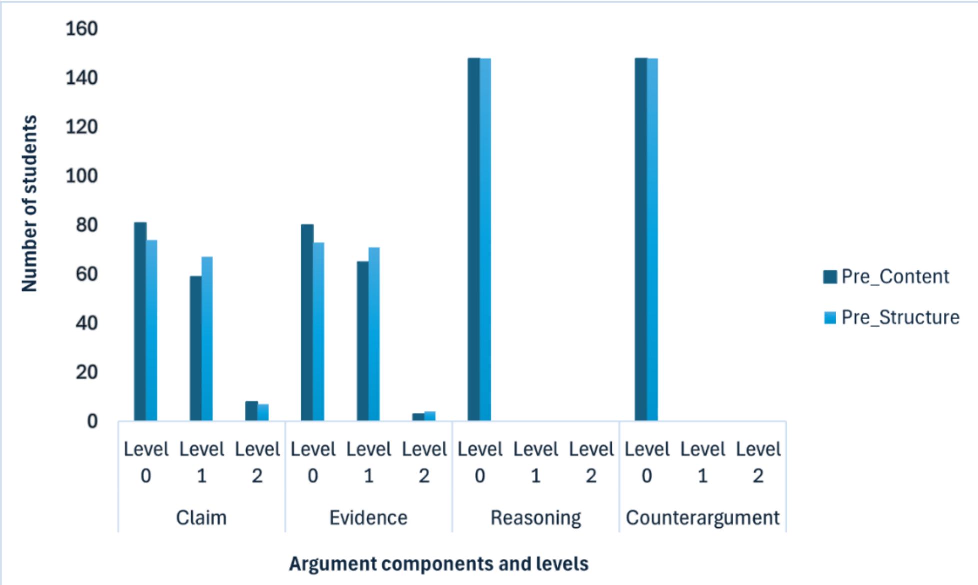
Examination of claims indicates that the majority of students struggled to formulate claims that were simultaneously contextually appropriate and structurally coherent. Specifically, 81 content responses were rated at Level 0, representing inappropriate or irrelevant claims, whereas 59 responses attained Level 1, and 7 responses reached Level 2. Structural ratings demonstrated a comparable pattern, with 74 responses at Level 0, 67 at Level 1, and 7 at Level 2.

Similarly, students exhibited considerable difficulty in integrating relevant evidence into their arguments. In the content dimension, 80 responses were rated at Level 0, 65 at Level 1, and 3 at Level 2. Structural ratings were comparable, with 73 responses at Level 0, 71 at Level 1, and 4 at Level 2.

With respect to reasoning, most student responses were at Level 0 in both content and structure, reflecting minimal engagement in explaining the link between evidence and claims. A small number of responses were classified at Level 1 (content: 2; structure: 3), representing partial inferential reasoning.

Engagement with counterarguments was similarly limited. Most responses were at Level 0 (content: 90; structure: 88), with a small proportion reaching Level 1 (content: 10; structure: 12), and none achieving Level 2, indicating minimal awareness of alternative perspectives and a lack of rebuttal strategies.

Figure 1
Content and structure analysis of students' worksheets pre-intervention



Improvement extent of students' SSA level about addressing urban heatwaves after the intervention

The educational intervention using the HEATWAVE CITY online serious game was associated with observable improvements in both the content and structure of students' arguments, across the two instruments of the research.

Questionnaire

Table 7 depicts the distribution of student responses across three performance levels for each item of the digital questionnaire, disaggregated by content and structural dimensions.

With respect to content, the intervention was associated with moderate enhancements, particularly in the claim and evidence components. In Q1, Level 2 responses increased from 23 to 48 for claim and from 23 to 50 for evidence. Comparable upward trends were observed in Q2, where Level 2 responses reached 52 for both components. In Q3, performance plateaued at 48, and in Q4, Level 2 responses declined to 30. These shifts were accompanied by a reduction in Level 0 responses, indicating that a greater proportion of students were able to construct content-relevant arguments. Nevertheless, overall gains in content remained uneven, with higher-level responses decreasing in later questions. In the domains of reasoning and counterargument, content-related progress was more limited. Although Level 0 responses decreased (e.g., from 100 to 75 in Q1), gains predominantly occurred at Level 1, and Level 2 responses remained negligible (Table 7) across all items post-intervention.

In contrast, structural improvements were more pronounced and broadly distributed across all components, suggesting a greater impact on students' ability to organize arguments coherently. For the claim component, Level 2 responses increased from 38 to 78 in Q1, from 60 to 83 in Q2, from 52 to 86 in Q3, and from 38 to 96 in Q4. evidence similarly exhibited steady improvement, with Level 2 responses rising from 55 to 82 in Q1 and reaching 100 in Q4. These developments reflect substantial progress in the logical structuring of claims and the effective integration of supporting evidence.

Although Level 2 responses were achieved in reasoning and counterargument post-intervention, noteworthy gains were observed from Level 0 to Level 1. For instance, Level 1 responses in reasoning increased from 48 to 73 in Q1 and from 24 to 80 in Q4. Counterargument followed a similar trajectory, with Level 1 responses increasing from 48 to 74 in Q1 and from 24 to 80 in Q4.

Table 7

Assessment of the content of students' responses to each questionnaire item pre- and post- intervention

Content

	Q1						Q2						Q3						Q4					
Components	Pre_0	Post_0	Pre_1	Post_1	Pre_2	Post_2	Pre_0	Post_0	Pre_1	Post_1	Pre_2	Post_2	Pre_0	Post_0	Pre_1	Post_1	Pre_2	Post_2	Pre_0	Post_0	Pre_1	Post_1	Pre_2	Post_2
Claim	75	30	50	65	23	53	80	28	32	60	36	60	90	25	35	65	23	58	100	20	38	71	10	57
Evidence	75	32	50	60	23	56	80	30	32	60	36	58	90	28	35	62	23	58	100	22	38	72	10	54
Reasoning	100	80	48	63	0	5	95	78	53	66	0	4	100	82	48	64	0	2	122	85	26	60	0	3
Counterargument	100	82	48	62	0	4	95	80	53	64	0	4	100	83	48	61	0	4	122	88	26	57	0	3

Structure

	Q1						Q2						Q3						Q4					
Components	Pre_0	Post_0	Pre_1	Post_1	Pre_2	Post_2	Pre_0	Post_0	Pre_1	Post_1	Pre_2	Post_2	Pre_0	Post_0	Pre_1	Post_1	Pre_2	Post_2	Pre_0	Post_0	Pre_1	Post_1	Pre_2	Post_2
Claim	55	22	35	48	58	78	50	18	38	47	60	83	70	16	26	46	52	86	70	12	40	66	38	70
Evidence	60	20	33	46	55	82	52	16	34	45	62	87	66	14	30	44	52	90	72	10	40	80	36	58
Reasoning	100	75	48	69	0	4	100	72	48	74	0	2	110	70	38	70	0	8	124	68	24	75	0	5
Counterargument	100	74	48	70	0	3	100	72	48	74	0	2	110	70	38	70	0	8	124	68	24	80	0	0

Low Medium High

Worksheets

As illustrated in Figure 2, structural improvements consistently exceeded gains in content across all components, particularly in claim and evidence, where structural clarity and coherence showed substantial enhancement.

Students' skills to construct claims demonstrated measurable improvement following the intervention. In terms of content, Level 1 responses increased from 59 to 90, and Level 2 responses rose from 7 to 22, while Level 0 responses declined from 81 to 36. Structural ratings followed a similar pattern, with Level 1 responses increasing from 62 to 96, Level 2 responses from 7 to 22, and Level 0 responses decreasing from 74 to 30.

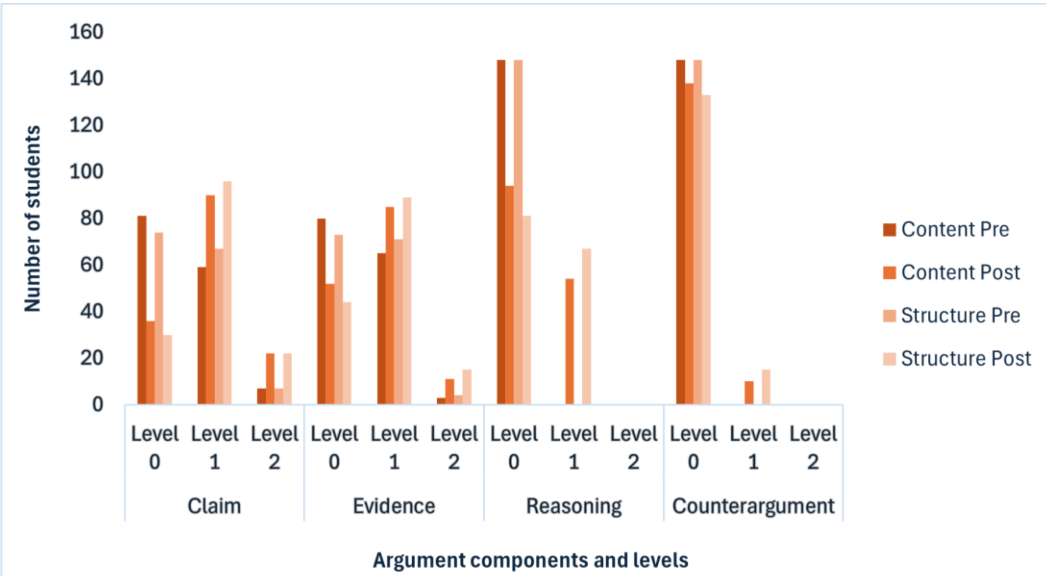
Performance in integrating supporting evidence also improved post-intervention. In the content dimension, Level 1 responses increased from 65 to 85, and Level 2 responses rose from 3 to 11, with Level 0 responses declining from 80 to 52. Structural improvements were somewhat more pronounced, with Level 1 responses increasing from 68 to 89, Level 2 from 2 to 15, and Level 0 responses decreasing from 73 to 44.

Prior to the intervention, nearly all student responses were classified at Level 0 in both content and structure reasoning. Post-intervention, 54 content responses and 67 structural responses reached Level 1, indicating emerging inferential reasoning skills. However, no responses achieved Level 2 in either dimension, highlighting ongoing challenges in developing higher-order reasoning to connect claims and evidence effectively.

Improvements in counterargumentation were modest. Content responses at Level 1 increased from 0 to 10, and structural responses from 0 to 15. No responses attained Level 2, reflecting limited engagement with alternative perspectives or rebuttal strategies.

Figure 2

Pre- and post-content and structure analysis of students' worksheets



A shift analysis was conducted to examine students' transitions across performance levels from pre- to post-intervention in both content and structure dimensions for the claim, evidence, reasoning, and counterargument components, using data from both the questionnaire and worksheets. This analysis aimed to capture the developmental trajectory of students' argumentative skills and to identify areas where growth was most pronounced.

For the questionnaire, shift matrices were aggregated across the four items for each component, separately for content (Figure 3) and structure (Figure 4). The total number of observations considered in the analysis was 592, corresponding to 148 students completing four items per component. Level 2 responses were excluded from transition counts, as students who initially performed at this level generally remained stable, indicating no further upward shift. Consequently, the matrices focus on transitions among students beginning at Levels 0 and 1.

Among students initially classified at Level 0, upward transitions were observed across all components in the content dimension. In the Claim component, 200 responses advanced by one level and 115 responses progressed two levels, while 30 responses remained at Level 0. Evidence demonstrated a similar pattern, with 200 responses shifting one level and 115 progressing two levels, and 30 remaining at the initial level. In Reasoning, 160 responses moved up one level and 14 progressed two levels, whereas 243 responses remained at Level 0. For Counterargument, 150 responses showed one-level advancement and 15 responses transitioned two levels, while 252 responses remained stable at Level 0. Students initially at Level 1 largely maintained their performance or demonstrated minor gains, with 75 Claim responses, 69 Evidence responses, 105 Reasoning responses, and 94 Counterargument responses advancing by one level, and no two-level transitions observed.

Shift patterns for structural components revealed pronounced upward transitions across all areas. In the claim component, 190 responses advanced by one level and 65 progressed two levels, while 90 responses remained at Level 0. Evidence exhibited a similar pattern, with 180 responses shifting one level and 65 progressing two levels, and 100 responses remaining at Level 0. Reasoning demonstrated substantial gains, with 261 responses moving up one level and 19 progressing two levels, whereas 137 responses remained stable at Level 0. Counterargument followed a comparable trajectory, with 255 responses advancing one level and 13 progressing two levels, while 149 responses remained at the initial level. Among students initially at Level 1, most maintained their performance or exhibited minor gains, with 82 Claim responses, 85 Evidence responses, 35 Reasoning responses, and 125 Counterargument responses improving by one level, and no two-level transitions observed.

Figure 3
Shift matrix of student transitions across content components of the questionnaire results

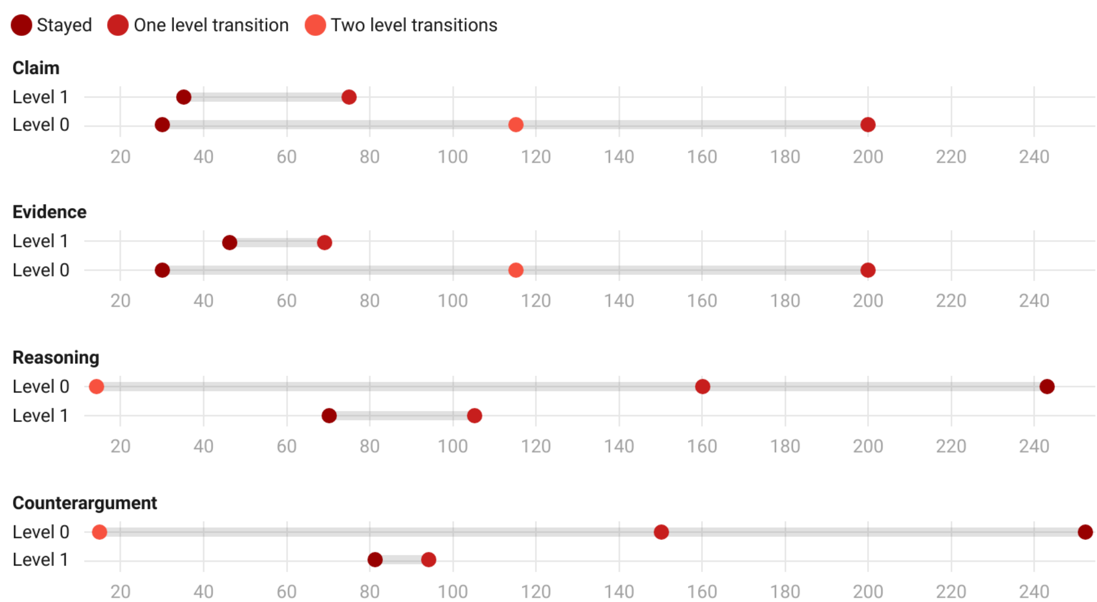
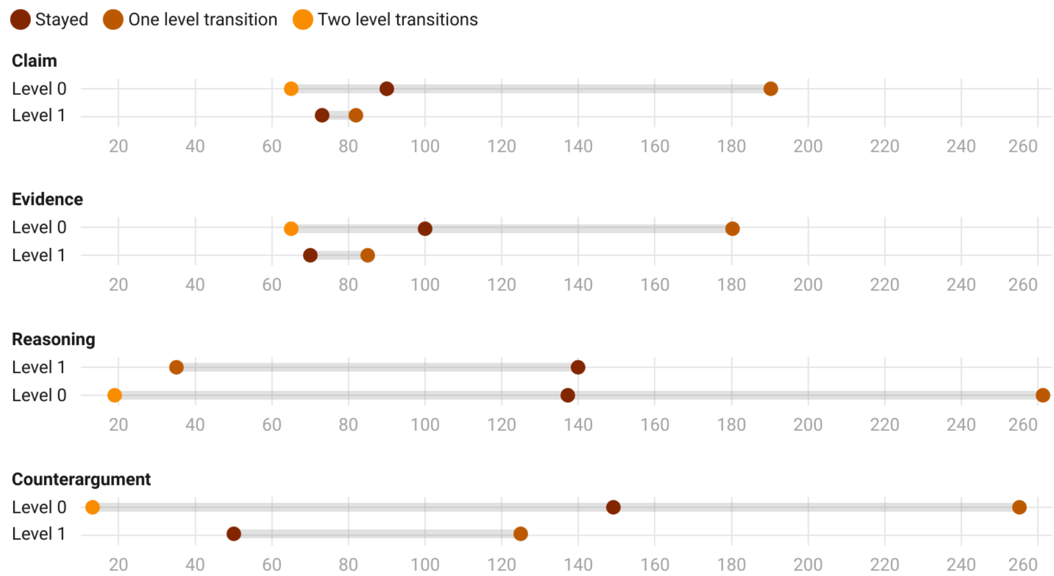


Figure 4
Shift matrix of student transitions across structure components for the questionnaire results



The shift analysis of worksheet data revealed distinct patterns of student transitions across both content and structure components.

In the content dimension (Figure 5), the Claim component exhibited moderate improvement. Among students initially at Level 0, 50 responses advanced by one level and 8 progressed two levels, while 20 responses remained at Level 0. Evidence showed a comparable pattern, with 40 Level 0 responses advancing by one level, 6 progressing two levels, and 32 remaining unchanged. In contrast, Reasoning demonstrated limited upward movement, with 54 responses moving from Level 0 to Level 1, while 94 responses remained at the initial level; no Level 2 responses were observed. Counterargument displayed minimal gains, with only 10 Level 0 responses advancing to Level 1 and 138 remaining at Level 0, and no two-level transitions recorded.

In the structural dimension (Figure 6), upward transitions were more pronounced, particularly in the Claim and Evidence components. For Claim, among students initially at Level 0, 58 responses advanced by one level and 10 progressed two levels, while 10 responses remained at the initial level. In Evidence, 48 Level 0 responses shifted one level and 12 progressed two levels, with 18 remaining unchanged. Reasoning demonstrated more limited progression, with 62 responses moving from Level 0 to Level 1, whereas 86 responses remained at Level 0, reflecting a pattern similar to that observed in content. In Counterargument, gains were minimal, with 15 Level 0 responses advancing to Level 1 and 133 remaining stable at Level 0; no two-level transitions were recorded.

Figure 5
Shift matrix of student transitions across content components of the worksheets results

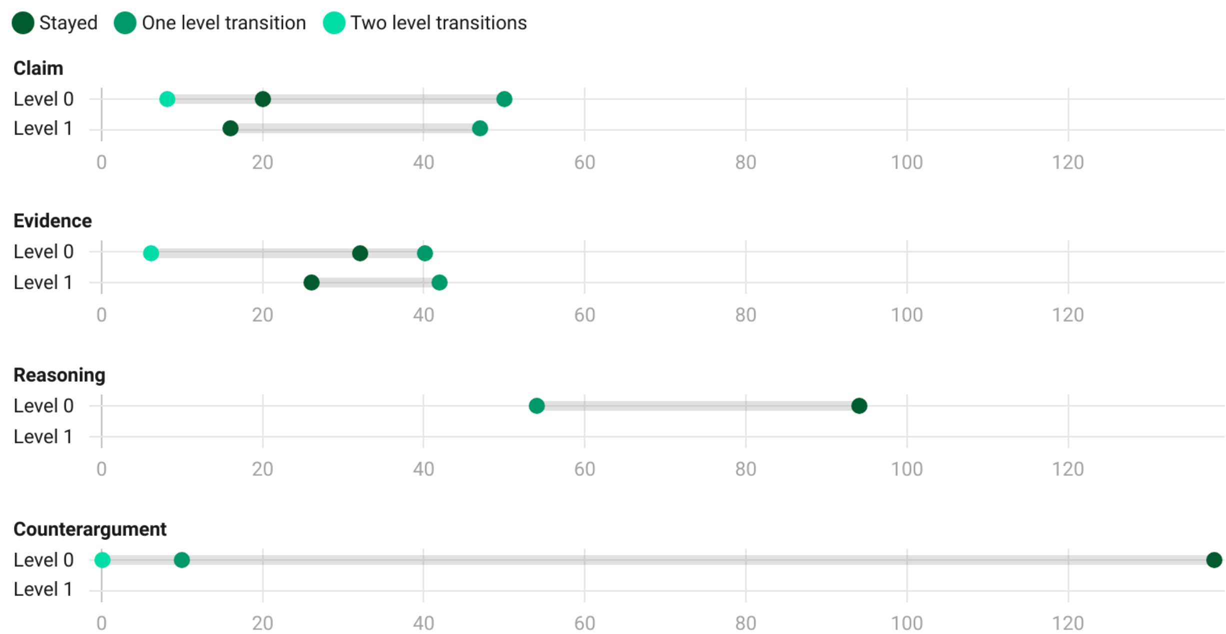
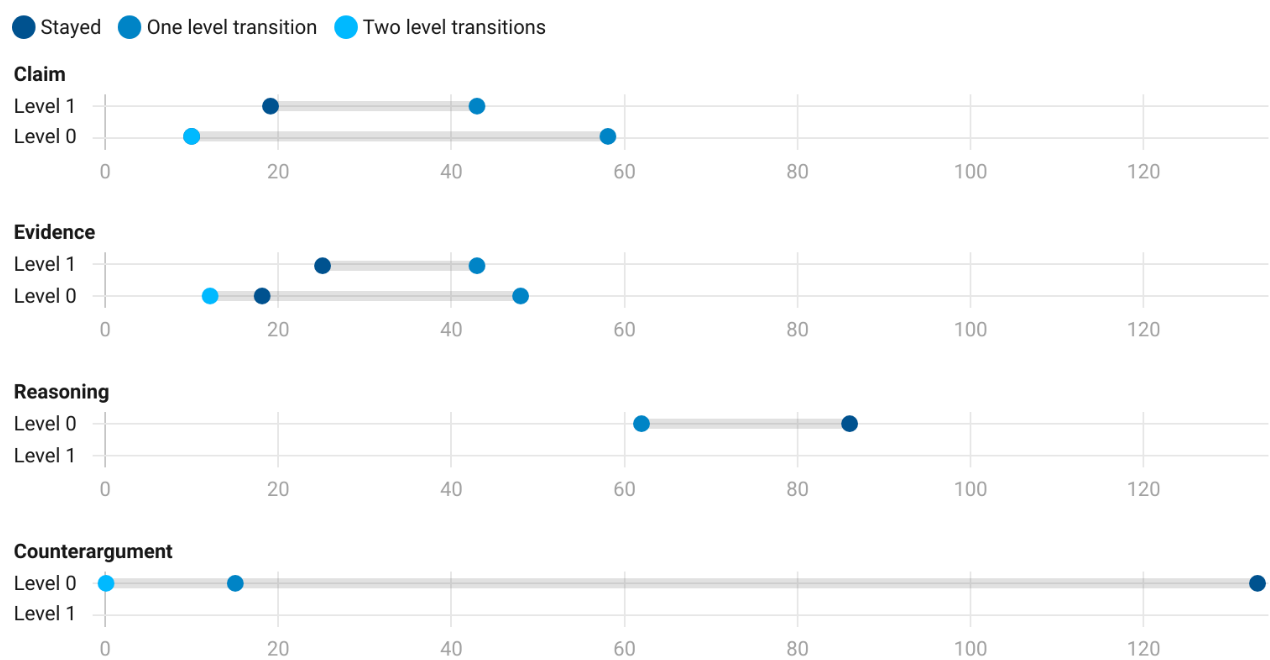


Figure 6
Shift matrix of student transitions across structure components of the worksheets results



Correlation of SSA level about combating urban heatwaves to the demographic variables

The analysis indicates that gender and school performance were systematically associated with students' ability to construct high-quality arguments, across both content and structure and across all components (claim, evidence, reasoning, and counterargument) (Table 8).

Male students had higher odds of achieving higher argumentation performance levels compared to female students pre-intervention (OR = 1.45, 95% CI [1.10, 1.91], $p = .008$). Students with higher school performance also had higher odds of achieving elevated argumentation levels compared to lower-performing peers (OR = 1.63, 95% CI [1.25, 2.13], $p < .001$). A significant interaction between gender and school performance (OR = 1.38, 95% CI [1.05, 1.82], $p = .020$) indicates that the influence of school performance on argumentation performance differed depending on gender.

These associations persisted following the intervention. Male students continued to have higher odds of achieving elevated argumentation performance levels (OR = 1.62, 95% CI [1.21, 2.17], $p = .002$), and high-performing students maintained higher odds relative to lower-performing peers (OR = 1.78, 95% CI [1.35, 2.34], $p < .001$). The gender × school performance interaction remained significant (OR = 1.42, 95% CI [1.09, 1.85], $p = .010$), demonstrating the consistent influence of demographic factors across time.

Table 8

Demographic associations pre- and post-intervention

Phase	Effect	Odds Ratio (OR)*	95% CI	p-value
Pre	Gender (Male vs Female)	1.45	1.10 – 1.91	.008
	School Performance (High vs Low)	1.63	1.25 – 2.13	<.001
	Gender × School Performance	1.38	1.05 – 1.82	.020
Post	Gender (Male vs Female)	1.62	1.21 – 2.17	.002
	School Performance (High vs Low)	1.78	1.35 – 2.34	<.001
	Gender × School Performance	1.42	1.09 – 1.85	.010

*Note:*Odds ratios were estimated using cumulative link mixed models (CLMM), accounting for repeated measures across items (Claim, Evidence, Reasoning, Counterargument) and classroom-level clustering. Holm-Bonferroni adjustments were applied for multiple comparisons.*

Discussion

The present study offers a nuanced exploration of how primary school students develop socioscientific argumentation (SSA) skills through an online intervention addressing urban heatwaves. As the boundaries between scientific understanding and societal concerns continue to blur, embedding SSA within science education emerges as an essential pathway for cultivating critical reasoning, ethical reflection, and epistemic awareness. These competencies enable learners to evaluate and synthesize evidence, to consider multiple perspectives, and to engage meaningfully with complex socio-environmental problems, an imperative for fostering civic scientific literacy.

Across the three phases of this study, a consistent pattern was observed: as task complexity increased, the quality of student argumentation declined, particularly in instances requiring interdisciplinary synthesis or higher-order evaluative reasoning. Pre-intervention results, which showed all participants performing at Level 0 in reasoning and counterargument, underscored a systemic limitation in students' initial ability to justify or refute claims. This finding resonates with the work of McNeill and Krajcik (2012), who similarly reported persistent difficulties among students in articulating robust justifications in scientific contexts.

Following the intervention, measurable improvements were evident in structural components of argumentation, notably in the formulation of claims and the use of evidence. Yet, these gains were less pronounced in reasoning and counterargumentation, revealing an important theoretical distinction between structural fluency and epistemic depth. Within Toulmin's Argument Pattern (1958), claims and data constitute the surface structure of argumentation, while warrants and rebuttals represent deeper epistemic processes that require an understanding of how scientific knowledge is justified and challenged. These higher-level processes, which integrate ethical, social, and evaluative dimensions, correspond to advanced stages of socioscientific reasoning (Sadler, 2009; Zeidler & Nichols, 2009).

This distinction was evident in the qualitative features of student responses. One student, for example, wrote:

> "We should plant more trees because trees make shade and keep the city cool."

While structurally coherent, this response lacks a warrant explaining why tree planting influences temperature regulation, relying instead on an intuitive causal link. By contrast, a more sophisticated response such as: "Planting trees helps cool cities because leaves release water vapor that lowers air temperature, but this works best if cities also reduce car use so pollution doesn't harm the trees" demonstrates a deeper integration of mechanistic and contextual reasoning. The difference between these two excerpts encapsulates the developmental trajectory identified in this study: structural elements improved with scaffolded instruction, yet epistemic reasoning and rebuttal remained comparatively underdeveloped. This aligns with previous research demonstrating that students acquire the formal conventions of argumentation more readily than the evaluative and dialectical thinking that underpin it (Erduran & Jiménez-Aleixandre, 2008; Osborne, 2010).

The observed decline in argumentative quality across the more complex tasks should not be interpreted simply as an outcome of cognitive fatigue, as such an explanation remains speculative. Rather, it likely reflects the absence of sustained scaffolding as cognitive demands increased. When learners are required to coordinate multiple layers of reasoning, scientific, ethical, and socio-economic, cognitive load intensifies, impeding their ability to construct coherent arguments (Sweller, 2011). This interpretation accords with evidence suggesting that repeated, structured opportunities to practice argumentation are essential for transferring reasoning competence across contexts (Kuhn & Udell, 2003; Berland & Reiser, 2009). Designing instructional cycles that integrate guided modeling, peer assessment, and reflective discussion could therefore strengthen the durability of these skills and deepen students' understanding of the epistemic functions of argumentation.

The socioscientific framing of urban heatwaves provided an interdisciplinary context linking scientific principles with ethical and environmental concerns. This contextualization enhanced student engagement by situating science within lived experience, making learning more relevant and motivating. Such findings are consistent with studies demonstrating that socioscientific issues (SSIs) stimulate students' moral reasoning and foster connections between science and citizenship (Sadler & Zeidler, 2005). At the same time, the analysis revealed statistically consistent gender differences across several components of SSA, with male students demonstrating comparatively stronger performance. These patterns should not be interpreted as indicative of inherent differences in reasoning ability, but rather as reflective of broader sociocultural dynamics that shape how students participate in and are recognized within science learning contexts. Norms governing classroom discourse, expectations around participation, and differential access to epistemic authority may influence whose contributions are taken up, elaborated, or sustained during argumentation-rich activities.

Prior research suggests that instructional settings emphasizing debate, justification, and critique can inadvertently privilege assertive communication styles and competitive forms of reasoning, practices that are often more closely aligned with socially reinforced norms of masculine participation (Ford & Wargo, 2012; Archer et al., 2015). From this perspective, the gender differences observed in this study draw attention to how instructional designs that appear pedagogically neutral may nonetheless reproduce existing inequities unless participation structures are intentionally designed to foreground inclusivity, dialogic engagement, and multiple ways of expressing scientific reasoning.

Nevertheless, gender and academic achievement emerged as moderating variables that influenced the intervention's effectiveness. Some students, particularly those with lower academic performance, showed less progress in reasoning. These findings underscore the need to conceptualize equity not only as an aspirational goal, but as an empirical concern that must be addressed through deliberate pedagogical design. Differentiated scaffolding, flexible pacing, and multimodal learning environments can support diverse learners and promote inclusive engagement with SSA.

Taken together, the findings reaffirm the necessity of an integrative approach to SSA instruction, one that balances structural organization with epistemic reasoning. Frameworks such as Claim-Evidence-Reasoning (CER) offer a coherent method for guiding students through the process of argument construction, while their alignment with Toulmin's model allows for explicit attention to the deeper reasoning components of warrants and rebuttals (Erduran, Simon, & Osborne, 2004; Zembal-Saul et al., 2019). Empirical evidence suggests that instructional tools including argumentation templates, guided inquiry tasks, and reflective rubrics can effectively scaffold learners' transition from procedural argumentation toward epistemic understanding (Berland & Reiser, 2011; Teknowijoyo et al., 2024). When situated within inquiry-based digital environments such as Heatwave city, these strategies appear particularly powerful in connecting argumentation with authentic, socially embedded scientific practices.

While the intervention produced measurable improvements in students' ability to formulate claims and integrate evidence, persistent challenges in reasoning and counterargumentation point to the need for continued instructional refinement. Longitudinal research is warranted to examine whether these skills are retained and generalized across different scientific and societal contexts. Moreover, qualitative analyses of classroom discourse and reflective writing could provide insight into how students negotiate evidence, values, and uncertainty, thereby illuminating the affective and moral dimensions of socioscientific reasoning (Zeidler & Sadler, 2008).

Ultimately, this study underscores the significance of sustained, interdisciplinary approaches that merge scientific understanding with ethical deliberation and strategic thinking. Developing scalable, evidence-based pedagogies that nurture such integration will be essential for preparing young learners to navigate the socioscientific challenges of the twenty-first century, reasoning not only with evidence, but with empathy, humility, and responsibility.

Conclusion

As attention increasingly turns to how science education can equip young learners to reason within complex socio-environmental realities, this study offers a timely contribution by framing socioscientific argumentation (SSA) not merely as a set of procedural skills, but as a multidimensional process of epistemic, ethical, and civic development. While the quantitative improvements observed in students' claim and evidence use are meaningful, their limited progress in reasoning and counterargumentation underscores that argumentation competence cannot be reduced to structural fluency alone. Rather, it must be understood as an evolving practice situated at the intersection of cognitive, moral, and contextual dimensions of learning.

The framing of this study within SSA foregrounds argumentation as a site of epistemic work—where learners negotiate not only what counts as valid evidence, but also how knowledge is justified and challenged within social contexts. In this respect, the findings invite a reconceptualization of argumentation pedagogy as a process of cultivating epistemic agency: the capacity to engage critically, reflectively, and responsibly with scientific and societal claims. This perspective extends beyond teaching students to assemble arguments correctly, urging educators to create conditions in which reasoning becomes both a cognitive and a moral endeavor.

In examining how students engaged with the socioscientific issue of urban heatwaves, the study illustrates how contextualized learning environments can bridge scientific understanding with lived experience, thereby rendering argumentation both authentic and consequential. Yet the disparities observed across gender and academic performance remind us that such opportunities are not equally accessible to all learners. The challenge, therefore, lies in designing pedagogies that democratize participation in reasoning—pedagogies that are inclusive, differentiated, and sustained across contexts.

Viewed collectively, these insights affirm that SSA serves as a powerful lens for examining and advancing the aims of contemporary science education. It provides a framework through which learning can be viewed not only as the acquisition of scientific literacy, but as the cultivation of epistemic and ethical dispositions necessary for civic participation in a scientifically complex world. The study thus contributes to an evolving understanding of how learners construct meaning at the boundaries of science and society, and how digital, inquiry-based interventions such as the Heatwave City online game, can support that process.

Limitations

While this study offers valuable insights into the development of SSA among primary students, several methodological limitations warrant careful consideration. First, the use of a 0–2 rubric scale to assess argumentation may have constrained the ability to detect subtle developmental changes. Such a compressed scale introduces potential floor and ceiling effects, particularly in pre- and post-intervention comparisons, and may have limited the sensitivity of our analyses. Future research would benefit from employing a more nuanced 0–4 or 0–5 scale, which could more accurately capture incremental growth in structural and epistemic components of argumentation.

Dependencies among responses from the same students, as well as potential clustering within classrooms or schools, were not explicitly accounted for in the analysis. Argumentation performance is likely influenced by shared instructional experiences, peer interactions, and classroom dynamics, which may introduce intra-class correlations. Ignoring these nested effects could affect the precision of statistical inferences. Future studies should consider hierarchical or multilevel modeling approaches to appropriately capture these dependencies and provide more robust estimates of intervention effects.

The present study relied primarily on pre- and post-test questionnaires and student worksheets, without integrating additional process-oriented data. Game logs, group discussion transcripts, and observational field notes could offer critical insights into the reasoning processes, collaborative dynamics, and real-time decision-making strategies that underlie observed outcomes. Including such process data in future research would enable a more comprehensive understanding of how SSA develops over time and in interactive contexts.

The study's context was limited to a single country, one grade level, and a single socio-scientific issue, urban heatwaves. These constraints restrict the generalizability of the findings, as developmental trajectories, instructional impacts, and socio-cultural factors may differ across grades, countries, or topic domains. The reliance on convenience sampling further limits the representativeness of the sample. Expanding future research to include broader, more diverse samples would deepen understanding of how educational interventions can foster sophisticated argumentation skills across varied student populations.

Furthermore, the investigation focused primarily on structural and epistemic dimensions of argumentation, without systematically examining a broader set of potential influencing factors. Cognitive capacities, socio-emotional dispositions, pedagogical approaches, classroom context, and cultural background are all likely to shape how students engage with socioscientific issues and develop argumentation skills. Incorporating these variables in future research would provide a more nuanced understanding of the mechanisms underlying change in SSA and allow for a more comprehensive analysis of student development.

Finally, the three-month post-test interval leaves the results open to potential uncontrolled influences and confounding factors. Students' engagement with science outside the intervention, exposure to related media, or classroom experiences could have affected the observed gains or declines in argumentation. Longer-term follow-ups, alongside careful documentation of concurrent experiences, are recommended to more accurately isolate the effects of targeted interventions.

Despite these limitations, the study provides meaningful evidence regarding the developmental trajectories of SSA in primary students and highlights critical areas for instructional refinement. Acknowledging these constraints offers a pathway for future research to enhance methodological rigor, expand generalizability, and deepen understanding of how young learners engage with complex socioscientific problems.

Acknowledgements

This research was funded by the Erasmus+ KA2 program (Project Number: 2022-1-EL01-KA220-SCH-000088530), with support from the National Agency for Erasmus+ (IKY). The funding agency had no involvement in the study's design, data collection, analysis, or interpretation.

References

Abell, S. K., & Roth, M. (1992). Constraints to teaching elementary science: A case study of a science-enthusiast student teacher. Science Education, 76(6), 581–595. https://doi.org/10.1002/sce.3730760603

Akerson, V. L., Townsend, J. S., Donnelly, L. A., Hanson, D. L., Tira, P., & White, O. L. (2009). Scientific Modeling for Inquiring Teachers Network (SMIT'N): The influence on elementary teachers' views of nature of science, inquiry, and modeling. Journal of Science Teacher Education, 20, 21–40. https://doi.org/10.1007/s10972-008-9116-5

Alegado, R. A., & Lewis, S. E. (2018). Claim-evidence-reasoning (CER) framework: Supporting students in constructing explanations in science. Pearson Allyn & Bacon.

Andrews, R. (2005). Argumentation and education: Theoretical foundations and practices. Springer.

Archer, L., DeWitt, J., Osborne, J., Dillon, J., Willis, B., & Wong, B. (2012). Science aspirations, capital, and family habitus: How families shape children's engagement and identification with science. American Educational Research Journal, 49(5), 881–908. https://doi.org/10.3102/0002831211433290

Berg, B. L. (2004). Qualitative research methods for the social sciences (5th ed.). Pearson Education.

Berland, L. K., & Reiser, B. J. (2009). Making sense of argumentation and explanation. Science Education, 93(1), 26–55. https://doi.org/10.1002/sce.20286

Berland, L. K., & Reiser, B. J. (2011). Classroom communities' adaptations of the practice of scientific argumentation. Science Education, 95(2), 191–216. https://doi.org/10.1002/sce.20420

Boston University. (2022). Differentiated instruction and equitable learning. https://www.bu.edu/education/differentiated-instruction

Capkinoglu, E., Yilmaz, S., & Leblebicioglu, G. (2020). Quality of argumentation by seventh graders in local socioscientific issues. Journal of Research in Science Teaching, 57(6), 827–855. https://doi.org/10.1002/tea.21609

Christenson, N., Chang Rundgren, S. N., & Zeidler, D. L. (2014). The relationship of discipline background to upper secondary students' argumentation on socioscientific issues. Research in Science Education, 44, 581–601. https://doi.org/10.1007/s11165-013-9394-6

Dawson, V., & Carson, K. (2020). Introducing argumentation about climate change socioscientific issues in a disadvantaged school. Research in Science Education, 50, 863–883. https://doi.org/10.1007/s11165-018-9715-x

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. (Eds.). (2007). Argumentation in science education: Perspectives from classroom-based research. Springer. https://doi.org/10.1007/978-1-4020-6670-2

Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's argument pattern for studying science discourse. Science Education, 88(6), 915–933. https://doi.org/10.1002/sce.20012

European Commission. (2019). The European Green Deal (COM(2019) 640 final). Brussels: European Commission. Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52019DC0640

Evagorou, M., & Osborne, J. (2013). Exploring young students' collaborative argumentation within a socioscientific issue. Journal of Research in Science Teaching, 50(2), 209–237. https://doi.org/10.1002/tea.21076

Ford, M. J., & Wargo, B. M. (2012). Dialogic framing of scientific content for conceptual and epistemic understanding. Science Education, 96(3), 369–391. https://doi.org/10.1002/sce.20482

Grooms, J., Sampson, V., & Golden, B. (2014). Comparing the effectiveness of verification and inquiry laboratories in supporting undergraduate science students in constructing arguments around socioscientific issues. International Journal of Science Education, 36(9), 1412–1433. https://doi.org/10.1080/09500693.2014.891160

Gül, M. D. (2020). Effectiveness of socio-scientific issues-based instruction on pre-service teachers' critical thinking skills [Master's thesis, Middle East Technical University]. National Thesis Center of Turkey.

Khishfe, R. (2023). Relationship between nature of science and argumentation: A follow-up study. International Journal of Science and Mathematics Education, 21, 1081–1102. https://doi.org/10.1007/s10763-022-10307-0

Kinslow, A. T., Sadler, T. D., & Nguyen, H. T. (2018). Socio-scientific reasoning and environmental literacy in a field-based ecology class. Environmental Education Research, 25(3), 388–410. https://doi.org/10.1080/13504622.2018.1442418

Kolstø, P. O. (2006). The sustainability and future of unrecognized quasi-states. Journal of Peace Research, 43(6), 723–740. https://doi.org/10.1177/0022343306068102

Kuhn, D., & Udell, W. (2003). The development of argument skills. Child Development, 74(5), 1245–1260. https://doi.org/10.1111/1467-8624.00504

Lawrence Hall of Science. (2024). Argumentation in science education: Strategies and tools. https://www.lawrencehallofscience.org/argumentation

McNeill, K. L., & Krajcik, J. (2012). Supporting grade 5–8 students in constructing explanations in science: The claim, evidence and reasoning framework for talk and writing. Pearson Allyn & Bacon.

McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. Journal of the Learning Sciences, 15(2), 153–191. https://doi.org/10.1207/s15327809jls1502_1

Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. Science, 328(5977), 463–466. https://doi.org/10.1126/science.1183944

Osborne, J., & Dillon, J. (2008). Science education in Europe: Critical reflections. The Nuffield Foundation. https://www.nuffieldfoundation.org/wp-content/uploads/2019/11/Sci_Ed_in_Europe_Report_Final.pdf

Sadler, T. D. (2009). Socioscientific argumentation: The effects of content knowledge and morality. Journal of Research in Science Teaching, 46(5), 531–554. https://doi.org/10.1002/tea.20326

Sadler, T. D., & Zeidler, D. L. (2005). The morality of socioscientific issues: Construal and resolution of genetic engineering dilemmas. Science Education, 89(1), 35–67. https://doi.org/10.1002/sce.20048

Sweller, J. (2011). Cognitive load theory. Psychology of Learning and Motivation, 55, 37–76. https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Toulmin, S. E. (1958). The uses of argument. Cambridge University Press.

United Nations. (2015). Transforming our world: The 2030 agenda for sustainable development (A/RES/70/1). New York, NY: United Nations. Retrieved from https://sdgs.un.org/2030agenda

Vicedo-Cabrera, A. M., Scovronick, N., Sera, F., Royé, D., Schneider, R., Tobias, A., … Gasparrini, A. (2021). The burden of heat-related mortality attributable to recent human-induced climate change. Nature Climate Change, 11(6), 492–500. https://doi.org/10.1038/s41558-021-01058-x

Wu, Y.-T., & Tsai, C.-C. (2011). The effects of different online searching activities on high school students' cognitive structures and informal reasoning regarding a socio-scientific issue. Research in Science Education, 41, 771–785. https://doi.org/10.1007/s11165-010-9189-y

Yamamoto, T., Kamiyama, S., Tanaka, T., & Yamaguchi, E. (2022). Primary school students' difficulties in writing arguments: Identifying challenges and opportunities for science teaching. Journal of Baltic Science Education, 21(3), 445–461.

Zeidler, D. L. (2014). Socioscientific issues as a curriculum emphasis: Theory, research, and practice. In N. G. Lederman & S. K. Abell (Eds.), Handbook of research on science education (Vol. 2, pp. 697–726). Routledge.

Zeidler, D. L., & Nichols, B. H. (2009). Socioscientific issues: Theory and practice. Journal of Elementary Science Education, 21, 49–58. https://doi.org/10.1007/BF03173684

Zeidler, D. L., & Sadler, T. D. (2008). The role of moral reasoning in argumentation: Conscience, character, and care. Journal of Research in Science Teaching, 45(7), 747–766. https://doi.org/10.1002/tea.20256

Zeidler, D. L., Sadler, T. D., Simmons, M. L., & Howes, E. V. (2005). Beyond STS: A research-based framework for socioscientific issues education. Science Education, 89(3), 357–377. https://doi.org/10.1002/sce.20048

Zembal-Saul, C., McNeill, K. L., & Herring, T. (2019). Argumentation and explanation in science education: The role of discourse. Journal of Research in Science Teaching, 56(4), 487–510. https://doi.org/10.1002/tea.21516

Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. Journal of Research in Science Teaching, 39(1), 35–62. https://doi.org/10.1002/tea.10008

Zölch, T., Maderspacher, J., Wamsler, C., & Pauleit, S. (2019). Using green infrastructure for urban climate-proofing: An evaluation of heat mitigation measures at the micro-scale. Urban Forestry & Urban Greening, 30, 305–316. https://doi.org/10.1016/j.ufug.2016.09.011

Corresponding Author Contact Information:

Author name:Christoforaki Maria

Department:Departement of Pedagogy and Primary Education

University, Country: National and Kapodistrian University of Athens

Email:mchrst@primedu.uoa.gr

ORCID: 0009-0003-3859-3470

Please Cite: Surname, . (year). Title of the paper. *The European Educational Researcher, 9*(1), 7-36. DOI: https://doi.org/10.31757/euer.912

Conflict of Interest: The authors declare no competing interests.

Data Availability Statement: This study did not generate publicly shareable datasets. Due to ethical and privacy considerations, data are not publicly available. Data access requests may be directed to the corresponding author and will be evaluated in accordance with institutional and participant confidentiality protocols.

Ethics Statement: This study was approved by the Ethics and Deontology Committee for Research of National and Kapodistrian University of Athens (Επιτροπή Ηθικής και Δεοντολογίας της Έρευνας – Ε.Η.Δ.Ε.) under Protocol Number 20566/5.3.2024. The Committee's decision number and date are 150/21.3.2024. Informed consent was obtained from all participants and/or their legal guardians prior to participation, in accordance with ethical standards.

Author Contributions: All authors contributed equally to the conception and design of the study, data collection, analysis and interpretation of results, and manuscript writing. All authors have read and approved the final manuscript.

Received: October 13, 2025 ▪ Accepted: January 12, 2026