Journal of Information Technology Education: Research

An Official Publication of the Informing Science Institute
InformingScience.org

JITEResearch.org

Volume 22, 2023

PROMOTING CRITICAL THINKING THROUGH ARGUMENT MAPPING: A LAB FOR UNDERGRADUATE STUDENTS

Francesca Crudele*	University of Padova, FISPPA- Department of Philosophy, Sociology, Pedagogy and Applied Psychology, Padova, Italy	francesca.crudele@phd.unipd.it
Juliana E. Raffaghelli	University of Padova, FISPPA- Department of Philosophy, Sociology, Pedagogy and Applied Psychology, Padova, Italy	juliana.raffaghelli@unipd.it

* Corresponding author

ABSTRACT

Aim/Purpose	In undergraduate training, helping students improve argumentative text comprehension (CoT) by identifying the elements of an argumentative text and critical thinking (CT) by reconstructing the meaning of the text and constructing their own reflections is relevant. Argumentative skills are essential on both the personal and professional levels.
Background	In recent decades, concern has developed over undergraduates' poor skills in reframing and articulating their thinking on a topic, which affects critical thinking and the ability to express unique perspectives. Customized interventions in higher education could develop argumentative abilities for professional and personal use. In this regard, argument maps (AM) could be a useful tool for the visualization of arguments. They provide logical relationships between statements to help understand the reasoning chain.
Methodology	Hybrid presence-distance research was conducted over four days. A quasi-experiment with one group and three tests – S1, S2, and S3 – was conducted.

Accepting Editor Peter Blakey | Received: August 17, 2023 | Revised: October 28, November 4, November 7, 2023 | Accepted: November 18, 2023.
Cite as: Crudele, F. & Raffaghelli, J. E. (2023). Promoting critical thinking through argument mapping: A lab for undergraduate students. *Journal of Information Technology Education: Research, 22,* 497-525.
https://doi.org/10.28945/5220

Contribution	Our study aims to investigate and enrich the research landscape, especially in the Italian university context, regarding the use of AM to support text comprehension and the development of argumentative skills.
Findings	Our preliminary descriptive analysis showed that AM improves students' CoT and CT proficiency levels. These abilities improved from the beginning to the end of the experiment. Inferential analysis showed a beneficial pathway inflection on final learning improvement. Early encouraging results suggest that AMs can improve argumentation comprehension, production, and critical thinking in teaching and learning.
Recommendations for Practitioners	The learner could better manage knowledge and understand different perspectives with AM usage rules.
Recommendations for Researchers	It is essential to remember that critical thinking is a multifaceted and complex concept. This article examined it as a proxy variable for text comprehension and argumentation skills. The length of exposure to the method, contexts and instruments (analog or digital), and age/education of participants must be considered when doing AM research.
Impact on Society	The method would improve awareness of shared ideas and make it easier to enrich and rethink one's thoughts on the topic.
Future Research	To study AM roles in diverse types of information, future research could incorporate quantitative and qualitative approaches. Cross-curricular learning for everyday life in digital and AI-driven environments, as well as text comprehension and critical thinking, could also be examined. Further research could cover other aspects of the topic of critical thinking.
Keywords	argument maps, comprehension text, critical thinking, argumentative skills, laboratory activity

INTRODUCTION

The act of communicating has always been a fundamental pillar of being human. Everyday life is full of moments when we interact with each other and engage in discussions (Canale et al., 2021). There is a chance of misunderstandings and disagreements because everything we think, read, or listen to affects how we interact with other people. Information collection and processing for conversation can be difficult (Colombo, 2018).

Despite being constantly immersed in communication and exchanges of information and opinions, in recent decades, there has been a pervasive preoccupation with the lack of comprehension of text and considerable difficulty in re-elaborating and expressing one's thoughts on a topic (Moretti, 2010). This problem is particularly relevant in the post-truth era (Hobbs, 2020). The changing digital spaces have led to conceptualizing our current time as the "post-digital" era. It assumes that the digital is not new and goes "beyond binaries such as online/offline, virtual/real, old/new media, digital/analog, technical/natural, biological/informational" (Macgilchrist, 2021, p. 660).

The enormous complexity faced by such a scenario prompts forceful thinking about media literacy and its promotion around the world (Hobbs et al., 2019). Indeed, with the rise of the Internet and digital culture, literacy has evolved to cover a variety of content. Many of these activate strong emotions in the user, simplify information, appeal to audience values, or attack opponents to influence attitudes and behaviors (Hobbs, 2020). In a sense, one could return to the term "propaganda" today. Consider how everyone's daily life, both online and offline, is affected by advertisements, sponsored content, as well as hoaxes, conspiracy theories, and pseudoscience (Hobbs, 2020; Hobbs et al., 2019).

Previous authors, such as Buckingham (2003, 2009), reflected on how the concept of media literacy was closely related to the idea of "critical awareness" and "democratic participation." They emphasized how the digitally literate individual could search efficiently, comparing a range of sources and selecting relevant and relevant documents (Buckingham, 2009). The push came for increasingly digitized information, requiring literacy that went beyond just functional computer use. We were moving toward the idea of evaluating, reflecting on, and critically using information to turn it into knowledge (Buckingham, 2006, 2015).

However, we had only just begun to realize that we needed to acquire new skills to deal with the highly dynamic, conflicting, and dangerous contexts (Hobbs et al., 2019) when the pandemic accelerated everything. Indeed, the pandemic has revealed ever-increasing tensions, and some of those changes are now irreversible (Coates et al., 2021). Understanding those changes is essential to understanding what is happening and what possible futures may emerge (Coates et al., 2021).

Throughout this context of continuous reconfiguration, recent literature still emphasizes the central role of promoting argumentative skills in a person's formal and informal education (Alotto, 2021; Colombo, 2018; Wambsganss et al., 2020). Kuhn (2010, 2018) spoke of a "double relationship" between argumentation as a critical thinking practice, leading to more argumentative thoughts, better decisions, and everyday problem-solving practices.

Today, the risk is not being able to grasp the point of view of others after reading an article, a post, or a contribution in a forum and failing to engage with each other constructively (Colombo, 2018). In this regard, in March 2010, the European Commission presented the New Strategy Europe 2020, which proposed, among the eight parameters, that the percentage of 15-year-olds with poor results in reading, math, and science should be lower than 15% (Commissione Europea, 2010). According to the monitoring reports of INVALSI (2015) on reading literacy in Italy, about 20% of the schooled 15-year-olds were not able to reach the second level in reading and comprehension on the Programme for International Student Assessment (PISA) scale (OCSE-PISA, 2015). Failing this level means being only able to recognize familiar words in simple text and establish meanings connected to the reading. This situation also remained unchanged in the PISA 2018 monitoring report (INVALSI, 2018) and some of the most recent results of OECD indicators (OECD, 2020, 2021).

These results find some continuity when we turn to the analysis of college students. The gap seems to transfer almost unchanged from the last years of high school to the first years of college. However, these are different moments in which the inability to understand what skills are needed to cope with the changing reality and especially how to acquire them persists.

In fact, the results of the annual surveys of the TECO (TEst on COmpetencies) project, launched in 2012 by ANVUR (National Agency for the Evaluation of the University System and Research), still constitute an integral part of the Student Self-Assessment, Periodic Evaluation, and Accreditation (AVA) system (ANVUR, 2021a). Through the construction of indicators that estimate the skill levels of university students, it aims to contribute to improving the quality of the educational process by activating internal mechanisms of self-evaluation within the academic world.

The most recent results of the TECO 2020/2021 tests, based on surveys of cross-curricular skills such as Literacy, Numeracy, Problem Solving and Civics in about 60 Italian universities, have revealed that in the very early years of university, statistically lower scores are found especially in Literacy and Numeracy (ANVUR, 2021b) than in older students. This points to a still firm difficulty in understanding, interpreting, evaluating texts and reflecting on the content and form of information of undergraduate students, especially in the early years. This is a difficulty that cannot be taken for granted, especially today.

In this context, the aim of this research is to analyze the impact of instruments that support argumentative skills, which are connected to critical thinking. We consider that appropriate instruments, both analogical and digital, could support the development of key skills to live in a postdigital society.

In the remainder of this paper, the reader will find specific sections with the following content: (a) background, introducing the importance of skills such as argumentative skills in post-digital reality and the methodology of AMs to support this; (b) method, where the research questions are stated; and (c) results, where we consider the findings in light of their relevance for future teaching and learning practices in contexts using multimedia, dynamic, and post-digital approaches. The discussion and conclusion sections illustrate the initial reconceptualization following the results obtained, with evidence of limitations and future design lines.

BACKGROUND

ARGUMENTATIVE SKILLS IN POST-DIGITALITY

Argumentative skills make it possible to go beyond simple argumentative performance, enabling critical thinking as the basis for citizenship in a democratic society (Iordanou & Rapanta, 2021). Knowing how to argue is not only an essential part of our daily lives but is also significant in developing collaboration and problem-solving skills (Wambsganss et al., 2020). The ability to argue is part of the ability to think critically and communicate to convince or persuade the reader (Iswati & Purwati, 2022).

Many authors have highlighted the central role of promoting argumentation skills in our formal education system. Most students, in fact, learn to argue through interactions with peers and/or teachers when they need valuable support in learning to proceed in argumentation (Wambsganss et al., 2020).

Argumentative texts have traditionally been seen as a tool for functional training in identifying diverse perspectives and their respective bases of support (Colombo, 2018). This textual typology, in fact, perfectly sums up the argumentative procedure – one argues a thesis about an argument and tries to provide supporting arguments or answer various objections to convince others of the goodness of the thesis (Lo Feudo, 2018; Özdemir, 2018). Training oneself to recognize and identify the six key elements – problem, thesis, reasons, antithesis, evidence, and conclusion – allows one to reconstruct the underlying thread of the argumentative proceeding (Özdemir, 2018). By learning the argument elements, the student will feel more solid even in the step-by-step creation of arguments (Özdemir, 2018). However, these components are not always made explicit. It proves necessary to perform actions and operations to be able to reconstruct what the author wants to communicate (Alotto, 2021, p. 17).

Many students still have difficulty understanding and analyzing the structure of argumentative text because it is a complicated task that requires skills that cannot be taken for granted (Alotto, 2021). The topics proposed in a text do not have a sequential nature *per se* (Alotto, 2021; Ganino, 2020). Very often, the key topic is surrounded by a variety of other superfluous propositions that interfere with comprehension and increase the so-called "cognitive load" for the reader (i.e., the effort associated with memory during cognitive activities such as learning or problem-solving) (Sweller, 1988, 2005). In contrast, information presented in a way that reduces this cognitive load would allow for increased learning and logical reasoning (Alotto, 2021; Ganino, 2020).

Argument Maps (AMs) can provide valuable support for this process. Unlike the more common mind maps and concept maps, they do not just connect ideas in a simple way; they also make complicated reasoning easier to understand by showing how different statements are related logically (Carrington et al., 2011; Lidåker, 2018). They are designed to enable the user to keep track of the chain of reasoning (Simari & Rahwan, 2009), understand it better, and be able to assess its correctness and acceptability (Alotto, 2021).

Clarity comes from the fact that students, through the map, are encouraged to reflect on the text read and reconstruct their existing knowledge. This process does not occur when students rework certain ideas forcibly proposed from outside (Mochizuki et al., 2019).

The rules for constructing AMs direct those who use them to apply a series of mental operations. Early in 1950, Monroe Beardsley ("Argument map," 2023) provided the first example of a text analyzed following a precise sequence of steps. Subsequently, more recent works by Harrell and Wetzel (2015) and Harrell (2022) have updated the procedure as follows:

- identify all the claims being made by the author;
- rewrite them as independent statements, eliminating non-essential words;
- identify which statements are premises, sub-conclusions, and the main conclusion;
- provide missing, implied conclusions, and implied premises (this is optional depending on the purpose of the argument map);
- put the statements into boxes and draw a line between any boxes that are linked;
- indicate support from the premise(s) to the (sub)conclusion with arrows.

By carrying out these cognitive operations, the construction of an AM becomes a moment of decoding and reconstruction of the argument (Alotto, 2021; Chounta et al., 2017; Harrell & Wetzel, 2015).

Diagramming, in short, forces one to dwell on what one reads. A range of skills of comprehension, analysis, representation, evaluation, and ultimately, the production of one's own reasoning are brought into play. Being able to grasp the meaning of what one reads, one can acquire the process of asking questions, examining data, and drawing personal considerations and conclusions from these (Davies, 2011, 2012; Harrell & Wetzel, 2015).

Argument Maps: Previous and Current Studies

Over time, this tool has been studied far and wide to test its potential. In his study on the effects of AMs, Christopher P. Dwyer (National University of Ireland) looked at how reading and then making structured maps to represent the topic could be useful in learning and assimilating activities in the classroom (Dwyer et al., 2010, 2013). In their study, Harrell and Wetzel (2015) emphasized the importance of developing critical thinking skills among college students, especially today. Hence, they experimented with the AMs technique, developing an entire "argumentative diagram curriculum." They studied the level of change in the development of different ideas students generated about the arguments of their own and others. Positive results were observed both in the development of general critical thinking skills and in the development of discipline-specific analytical skills.

In a study done at Princeton University, van der Brugge (2018) of Melbourne University found that teaching philosophy with AMs helped students improve at critical thinking and reworking. More recent studies by Fan and Chen (2021) from the Department of Computer Science and Information Engineering at Taiwan University uncovered how a computer-assisted AM and argumentative essay-writing system supported students in learning argumentation structures and improving their argumentation skills.

Considering the previous background, the present study aims to analyze the impact of AMs on students' level of text comprehension and explore the liaisons with critical thinking skills. Moreover, the study adopts a hybrid approach to learning, where presence and distance learning are blended, in an attempt to understand whether this environment is effective in promoting more argumentative skills. Indeed, the shift to emergency online learning due to the closure of so many campuses and universities was the first push. But students' maturity in engaging in hybrid learning is still a work in progress (Coates et al., 2021).

In the Italian context, we observed a dearth of empirical research. Despite the existence of tools created and used in specific cases (cf. Alotto's (2021) research previously mentioned), the research field connected to the impact on argumentative skills and critical thinking has not been followed up according to our screening of the literature in Italian.

Indeed, to support the present investigation, the authors conducted a prior analysis based on a systematic literature review (Crudele & Raffaghelli, in press). As a first action, a screening of Italian

journals (listed in ANVUR, a selection of best quality journals at Italian level, https://www.anvur.it/wp-content/uploads/2023/09/Elenco-riviste-classeA_Area11_05092023.pdf) was conducted using the queries "argument maps" and "argument maps 'AND' critical thinking" but no case studies were found in Italy about the subject matter. Previous thinking was somewhat confirmed.

Instead, the study intensively documented the use of AM methods to develop argumentative understanding, writing skills, and critical thinking internationally. Nonetheless, AM usage was relatively new in teaching and learning digital contexts. Most studies adopted an "analog" setting. For example, in Kaeppel's (2021) study, a course with AMs was conducted in an undergraduate philosophy course to provide support for the development of critical discernment of meanings and different points of view. The setting was purely analog, with analog argumentative texts and maps. A deep connection between AM and critical thinking was found here. Another example of using an analog setting is the study by Malmir and Khosravi (2018). This study analyzed the effect of teaching AMs technique on the writing achievement of Iranian EFL (English as a foreign language) learners. The results of the data analysis revealed that AM strategies had a significant impact on improving expository and descriptive writing tasks.

The digital and hybrid learning approach is in the early stages of development. For example, the study by Kabataş MemiŞ and Karakuş (2021) presented both digital argumentative text and analog AMs. The experiments conducted tested the use of AMs, respectively, to support students' reading and comprehension skills of argumentative texts in relation to improving problem-solving skills. Another example is Eftekhari and Sotoudehnama's (2018) study with digitized mapping tools. The aim was to increase the level of text comprehension related to the development of more transversal skills, such as memory and memorization of the topic structure. In both studies, the results demonstrate a positive effect of AMs on learning. However, the results cannot be considered fully generalizable and relatively contextual (the first study was conducted in a Turkish secondary school context, and the second in a group of Iranian EFL undergraduates). Stemming from the previously mentioned international research landscape, our study in the specific context of Italian undergraduate education finds justification.

Although our study might not be deemed generalizable given the quasi-experimental nature, it sheds light on possible nuances and specific problems in a local context. It lays the groundwork for other comparative studies, paving the way for comprehensive reflection on the effectiveness of AMs in promoting argumentative skills (comprehension and production of argumentative texts) and, not least, critical thinking in an age of post-truth and post-digital.

MATERIALS AND METHODS

AIM AND RESEARCH QUESTIONS

The research activity aimed to probe: (1) the extent to which AMs, delivered in a hybrid course, aid students in increasing their level of text comprehension and critical thinking skills, and (2) the extent to which improving these argumentative skills connect with self-evaluative perception and final performance.

Based on these premises, we formulated one primary research question (PRQ) and three subsidiary research questions:

> **PRQ** (Primary research question): Is there a relationship between AMs intervention and increased text comprehension related to critical thinking and final performance?

Starting from this general question, more specific research questions (SRQ) were traced:

> **SRQ 1.1:** Does the training on AM usage support the increase in comprehension of an argumentative text?

This question aims to explore to what extent the technique of AMs correlates with the increase in undergraduate students' understanding of a given text in the context of hybrid learning. The AMs, embedded into a hybrid course configure the treatment (or independent variable). Students' understanding of a text configures the response explored (or dependent variable). Specifically, students' understanding is analyzed, and the correct identification of the text's structural elements is considered, as explained before.

SRQ 1.2: Does the training on AMs' usage develop critical thinking?

This question aims to explore to what extent the AM technique is related to the development of undergraduate students' level of critical thinking in the context of hybrid learning. The AMs, embedded into a hybrid course, configure the treatment (or independent variable). The level of critical thinking related to text comprehension configures the explored response (or dependent variable). Specifically, students' critical thinking is analyzed and considered as the correct reconstruction of the author's opinion of the text and as the construction of their own opinion about it.

SRQ 1.3: Does the increased level of argumentative ability relate to the final performance in terms of perception and actual output?

This question aims to explore to which extent the technique of AMs is related to the improvement of undergraduate students' argumentative ability and their final performance. Here the level of argumentative strategies in free writing on forums (or dependent variable) and the final grade (or second dependent variable) configure the explored responses. Specifically: (1) students' argumentative skills are analyzed as the detection of the use of argumentative strategies in writing posts on the general forum; (2) self-perception of improvement is analyzed in terms of student's self-evaluation; and (3) students' final performance is analyzed as the final grade at the end of the course.

Sample

The target sample numbered 103 female students, aged 18 to 40, attending the second year of the three-year course in Design, Documentation, and Evaluation for Early Childhood in the University of Padua's Bachelor of Science in Education and Training (L-19), based in Rovigo.

Research Design

The research was conducted through a hybrid approach (Presence-Distance) over four days of activities (Figure 1 and Table 1). A quasi-experimental design with a single-group experimental plan was chosen. The reason for this choice is that in an ecological context such as university and university courses, this type of study is considered preliminary and the first stage of a larger study. In addition, sometimes the choice depends on general planning circumstances and other issues not directly related to teaching. The activity proposed was a "Lab" as a space where the students can actively engage with resources, activities and the teaching staff. The topic offered also proposed a degree of novelty (technologies and AI in early education and care) in order to expose the students to aspects that generated a sense of puzzlement as well as curiosity. The research followed the general ethical instructions and guidance provided by the FISPPA department's ethics committee guidelines. In addition, for the improvement of practices, pedagogical practices followed the continuous improvement of the T4L program.

We created a hybrid setting that combined in-presence moments and asynchronous teaching methods. The hybridity consisted of digitally usable argumentative texts, digital information sources such as tutorials that replaced the in-person lecture, and a series of activities to be done on the course's Moodle platform. Our intent was to see if the hybrid study could prove as functional and/or effective as an analog study.

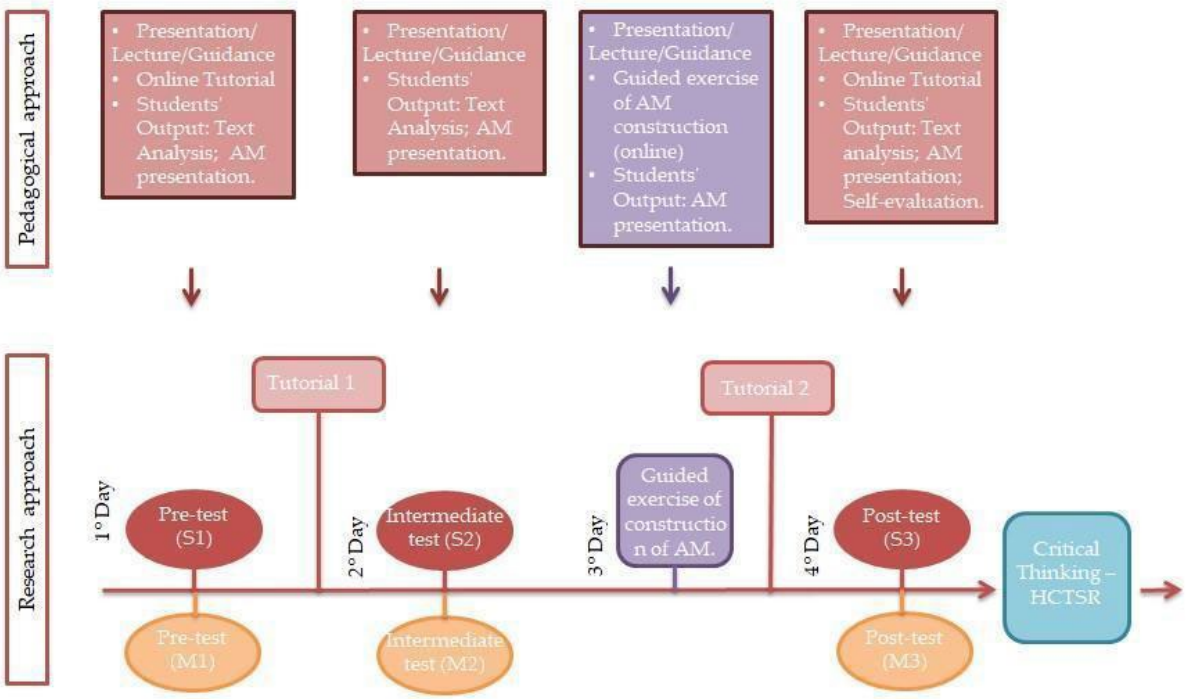


Figure 1. Outline research design

Table 1. Method instruments and procedures

Teaching activities	Moments	Data collection tools	Data analysis tools	Expected results per SRQ
Introduction to the topic	At the beginning of each step	Exposition to a topic triggering participants' response	Participation	The focus of Texts' Analysis and AM construction
Comprehension of texts data	Pre-test (S1)	The semi-structured survey instrument with closed stimulus questions.	Checklist of argumentative correctness.	SRQ 1.1. Increased comprehension of argumentative text, in terms of identifying structural elements.
	Intermediate test (S2)			
	Post-test (S3)			
AMs construction	Pre-test (M1)	Graphic elaboration of an AM.	Checklist of correct map construction.	SRQ 1.1. Increased AMs construction skills, in terms of identifying and juxtaposing components in space.
	Intermediate test (M2)			
	Post-test (M3)			

Teaching activities	Moments	Data collection tools	Data analysis tools	Expected results per SRQ
Critical Thinking data	Pre-test (CT1)	Based on text comprehension data	An adapted version of the Holistic Critical Thinking Scoring Rubric (HCTSR).	SRQ 1.2. Increased level of critical thinking, in terms of reconstructing the meaning of the text and constructing one's own thinking.
	Intermediate test (CT2)			
	Post-test (CT3)			

A single-group quasi-experiment was conducted with three testing moments: pre-test (S1), mid-test (S2), and post-test (S3). To collect data about argument comprehension, in terms of identifying the basic structural components of an argumentative text, a semi-structured survey instrument was created with closed-stimulus questions (Appendix A and Appendix B). The tool presented precise tasks to which the student had to respond with short, placeable answers within a limited range of possibilities. The correctness of the answers would then refer to a predetermined pattern of correct answers: a checklist was constructed with scores of 0 - 0.5 - 1, respectively, for an undetected, partially identified and identified component, for a total of 7 points.

Specifically, students were asked to read an argumentative text based on a course topic (technologies and childhood), so as to preserve familiarity and then to fill out the questionnaire, reporting the elements of the argumentation (problem, thesis, supporting arguments, objections and conclusion). Finally, they were to construct an AM (M1, M2, M3), bringing together the components of the text they read and reconstructing its meaning. A checklist with scores from 0 to 7 was created to evaluate the construction of the maps. The completeness of the maps was measured by how many structural elements were identified and juxtaposed during construction.

Based on the data collected about the comprehension of texts, an adapted version of the Holistic Critical Thinking Scoring Rubric, the HCTSR (Facione & Facione, 2014), a holistic rubric designed to assess the level of critical thinking qualitatively, was employed (Appendix C). The original version of the rubric included a grid of four mastery levels and some input for rethinking action per level (Facione & Facione, 2014). After about three meetings among the paper's discussants, it was decided to combine the rubric's mastery levels with four specific categories, arriving at the most correct correlation between the rubric's indices and the five key components-texts to be identified.

The rubric maintained a grid of four levels of mastery (strong, acceptable, weak, and absent), combined with four investigated categories obtained from the text comprehension test questions:

- Identify important information.
- Identify arguments and counter-arguments or alternative points of view.
- Conclude and explain the reasons.
- Understand and modify one's opinion based on evidence.

The instrument was not directly administered to students but was used by the researchers at three separate times: pre (CT1), intermediate (CT2) and post (CT3) times. Students were then asked to make their opinions about the tools used explicitly through posts on forums.

In this study, we assessed the reliability and internal consistency of the rubric and its components using Cronbach's Alpha analysis. Measurement was sought first by single dimension and then in total. Initially, we calculated Cronbach's Alpha separately for each dimension and then for the overall set. Overall, the results indicated moderate internal consistency among the dimension items (pre = 0.71;

int = 0.4; post = 0.46). However, when examining the overall Cronbach's Alpha, we observed a relatively high value (raw alpha = 0.71), suggesting satisfactory internal consistency among the items. Furthermore, the standardized alpha value (std. alpha = 0.71) indicates that the items contribute consistently to the overall internal consistency. Overall, these findings demonstrate robust internal consistency among the items comprising the rubric used in our study.

To answer SRQ 1.3. data were prepared and collected about different tests to capture a correlation between the variables analyzed: the structured knowledge test, the comprehension of text; the unstructured test, such as the free writing forum; the self-assessment moment; and the final performance return moment, with the final grade (Table 2).

From there, a multiple linear regression was conducted to estimate the relationship between two or more independent variables and one dependent variable. Two regression models were built with different predictors, investigating final performance and the critical thinking construct as the response variable, respectively.

To build the models, the hierarchical regression method was considered. Predictors are selected based on previous work and entered into the model. Once entered, the experimenter can delete and/or add any new predictors to the model in a stepwise manner (Field et al., 2012, p. 264). The regression models considered, as we will see later, were analyzed and enriched in this manner.

Table 2. Analysis phases

Phases	Variables	Data collection tools	Analysis methodologies and tools	Adopted Technologies
Descriptive Statistics Phase	Text Comprehension	Check-list	One-way repeated-measures ANOVA test	Moodle platform (collection data) R studio (analysis data)
	AMs Construction	Check-list		
	Critical thinking	HCTSR		
Preparatory Phase	Argumentative Skills in Forum	Check-list	Spearman Correlation	Moodle platform (collection data) R studio (analysis data)
	Self-assessment	Structured Questionnaire		
	Final Performance	Final grade		
Regression Phase	Final Performance	Final grade	Multiple Linear Regression	Moodle platform (collection data) R studio (analysis data)
	Critical Thinking	HCTSR		

The appendices have been organized to present the instruments used in data collection more specifically. Appendix A presents the semi-structured survey instrument with the seven closed stimulus questions. Appendix B is the checklist for scoring the correct text comprehension answers, as explained before. Appendix C presents the adapted version of HCTSR, with its four mastery levels related to the four indicators obtained from the comprehension questions.

RESULTS

In the following sections, the findings are presented in response to the proposed research questions.

SRQ 1.1: DOES THE TRAINING ON AMS' USAGE SUPPORT THE INCREASE IN COMPREHENSION OF AN ARGUMENTATIVE TEXT?

To answer this research question, the data collected about the correct understanding of the argumentative text and the respective construction of an AM were analyzed. The first scores obtained concern the number of correctly identified structural components of the text. Then, based on the identified elements, the correct construction of AMs was analyzed. The boxplots in Figure 2 introduce the distribution and variability of the first two variables analyzed.

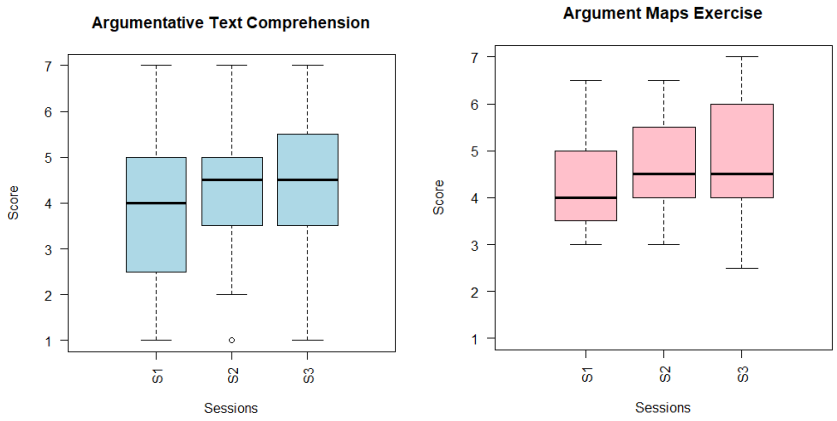


Figure 2. Boxplot

Regarding the ability to correctly identify and report the structural components of an argumentative text, after the course with AMs, students are facilitated in reporting the correct components (S3. M=4.48; SD=1.34) (Table 3). Except for S2 (S2. Me=4.500; M=4.231; IQR=2.125), where there was greater variance, the median was close to the mean in general. We found stability in the impact of experimentation (S1. Me=4,000; M=3,946; IQR=2,250; S3. Me=4,500; M=4,480; IQR=2000).

Table 3. Descriptive statistics - text comprehension test

Situation	Mean	Std. Dev.	Min	Q1	Median	Q3	Max	Skewness	Kurtosis
Students' comprehension of the text									
S1	3.95	1.43	1.00	2.50	4.00	5.00	7.00	-0.04	-0.67
S2	4.23	1.35	1.00	3.25	4.50	5.50	7.00	-0.33	-0.43
S3	4.48	1.34	1.00	3.50	4.50	5.50	7.00	-0.34	-0.46

The descriptive statistics table shows a rise in text comprehension scores from start to finish. Inferential tests determined the significance of teaching activities as can be seen in the Open Data (Crudele & Raffaghelli, 2023). Inferential tests revealed a moderately significant difference in text comprehension, with a post-hoc test showing a significant difference between pre-test and post-test scores (See Appendix D for details.). Ultimately, the effect size confirmed a moderate effect between pre-test and post-test. This result seems to indicate that AMs facilitate better identification of structural elements in argumentative texts and subsequent comprehension.

Continuing with the data on argument map construction, on average, there was a slight improvement from the initial phase (M3. M=4.68; SD=1.22). In this case, however, more divergent values were found between the median and mean (M1. Mdn=4.000; M=4.462; IQR=2.000; M2. Mdn=4,500; M=4,583; IQR=1,500; M3. Mdn=4,500; M=4,675; IQR=2,000). This suggested some variability and possible diversification concerning the performance of students (Table 4).

Table 4. Descriptive statistics - argumentative map construction

Situation	Mean	Std. Dev.	Min	Q1	Median	Q3	Max.	Skewness	Kurtosis
Completeness of maps constructed by students									
M1	4.46	1.10	2.50	3.50	4.00	5.50	7.00	0.46	-0.97
M2	4.58	0.95	3.00	4.00	4.50	5.50	6.50	0.21	-0.96
M3	4.68	1.22	2.00	4.00	4.50	6.00	7.00	0.09	-0.91

The descriptive statistics again revealed an improvement at the final phase of experimentation. However, inferential tests revealed a widespread low significance, with a post-hoc test showing no significant difference between testing moments (Appendix D). Ultimately, a small effect size was found and confirmed in all three moments. The average results, therefore, seem to have improved at the final phase. However, the low significance, probably caused by the greater dispersion of students' performance, i.e., performance that is too different from each other, suggests the need to reflect on the greater complexity of the map-making skill.

SRQ 1.2: Does the Training on AMs' Usage Develop Critical Thinking?

An adapted version of the HCTSR holistic rubric was used to answer this research question. Data were collected on students' critical thinking in relation to text comprehension responses. The final result took into account the reconstruction of the meaning of the text in terms of how many structural elements of the argumentative text were identified and the level of structuring one's opinion on the topic covered. The boxplot in Figure 3 shows an initial visualization of the variable's trend.

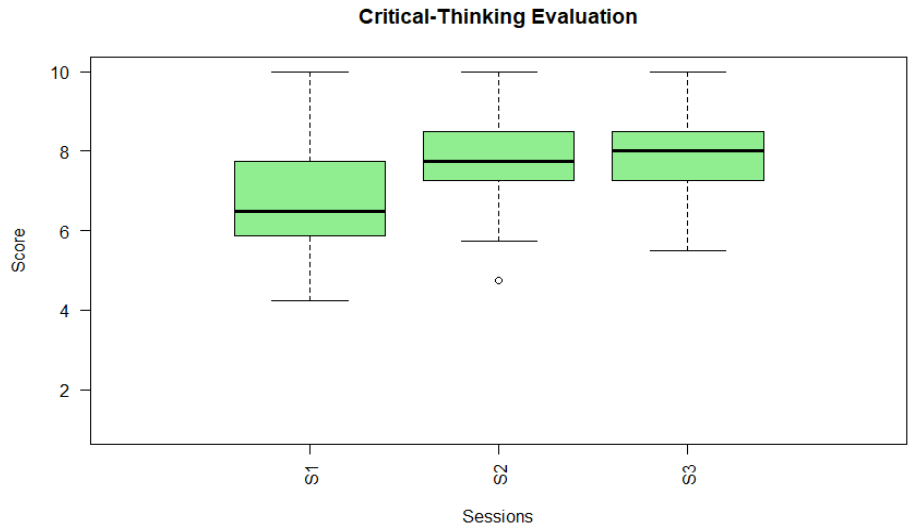


Figure 3. Boxplot

Students' critical thinking improved after the AM course, with steady improvement from the initial phase. Stability was observed in detecting the experimentation's impact on understanding opinions on topics (Table 5).

After the course with AMs, an increase in students' critical thinking was detected (CT3. M=8.00; SD=1.03). From the initial moment (CT1), there was a steady improvement with low, if any, deviation (CT1. Mdn=6.500; M=6.660; IQR=2.250; CT2. Mdn=7,750; M=7,780; IQR=1,250; CT3. Mdn=8,000; M=8,000; IQR=1,250). It was then possible to read stability in detecting the impact of experimentation on the skills of understanding one's own and others' opinions on a topic (Table 5).

Table 5. Descriptive statistics - level of critical thinking

	Mean	Std. Dev.	Min	Q1	Median	Q3	Max	Skewness	Kurtosis
Level of Critical Thinking with HCTSR Rubric									
CT 1	6.66	1.44	4.00	5.50	6.50	7.75	10.00	0.28	-0.62
CT 2	7.78	1.07	4.25	7.25	7.75	8.50	10.00	-0.33	0.45
CT 3	8.00	1.03	5.50	7.25	8.00	8.50	10.00	0.02	-0.51

The descriptive statistics revealed an improvement at the final phase of experimentation. Inferential tests and post-hoc found a significant difference in critical thinking data between two testing times, CT1-CT2 and CT1-CT3 (Appendix D, copy of Table 5). This indicated an increasing improvement in understanding argumentative text opinions and formulating one's own thinking. The self-perception of substantial improvement in female students from the early stages of the experiment suggests a high subjective impact.

SRQ 1.3: Does the Increased Level of Argumentative Ability Relate to the Final Performance in Terms of Perception and Actual Output?

To answer this research question, students' contributions on the forum, self-assessment data, and final performance results, were collected.

In the preparatory phase, which can be seen below, the correlation between the variables was traced. It was used to establish the strength and direction of the relationship between the laboratory variables and the continuous learning variables. The data collected during the intervention and the use of AMs were correlated with data on: (1) students' argumentative skills in terms of more structured written posts in the forum; (2) self-perception of improvement; and (3) final performance.

The second stage, the regression stage, was conducted to investigate further the relationship between a construct (dependent variable) and multiple independent variables. It was interesting to investigate which independent variable influenced the response of the model the most.

SRQ 1.3.1: Preparatory phase

The study investigated three variables: argumentative ability related to free production on the forum, subjective perception, and final performance. In the APA-style table (Appendix E), a correlation matrix between all variables in the model is shown.

The results showed no significant Spearman correlation between Forum_Post, the final detection of argumentative skills in the forum, and Post_CriThink, the final level of critical thinking. However, a positive correlation was observed with Map3b_Lab, the final skill of constructing AMs, and a positive moderating correlation with PostTest_Lab, the final moment of text comprehension (Appendix D).

The second aspect, students' subjective perception, showed a negative correlation between Autoval_Post, PostTest_Lab and Post_CriThink.

Regarding final performance, results revealed a significant positive correlation with Map3b_Lab and a moderate positive correlation with the variables PostTest_Lab and Post_CriThink. A heatmap graph revealed a deep positive correlation between PreTest_Lab and Pre_Crithink, and PostTest_Lab and Post_Crithink. Excessive linearity emerges between the two predictors, perhaps due to the strong interconnectedness of the data collection and analysis tools.

SRQ 1.3.2: Regression phase

The first regression model was constructed (cf. open data), including PreTest_Lab, PostTest_Lab, Map1b_Lab, and Map3b_Lab as explanatory variables and Grade_Fin as the response variable (Appendix D). The model showed a good fit to the data. The explanatory variable Map3b_Lab contributed significantly to the model, but only 14% of the variance was explained.

The second regression model was constructed, including PreTest_Lab, PostTest_Lab, Map3b_Lab and Forum_Post as explanatory variables and the Post_Crithink as a response variable. No normal distribution was checked. Therefore, it was adjusted by removing predictors (Forum_Post) and investigating the influence of the explanatory variables PreTest_Lab, PostTest_Lab, Map1b_Lab and Map3b_Lab. The regression equation was significant, explaining about 61% of the variance in Post_CriThink scores. The PostTest_Lab predictor contributed significantly to the model.

DISCUSSION AND CONCLUSIONS

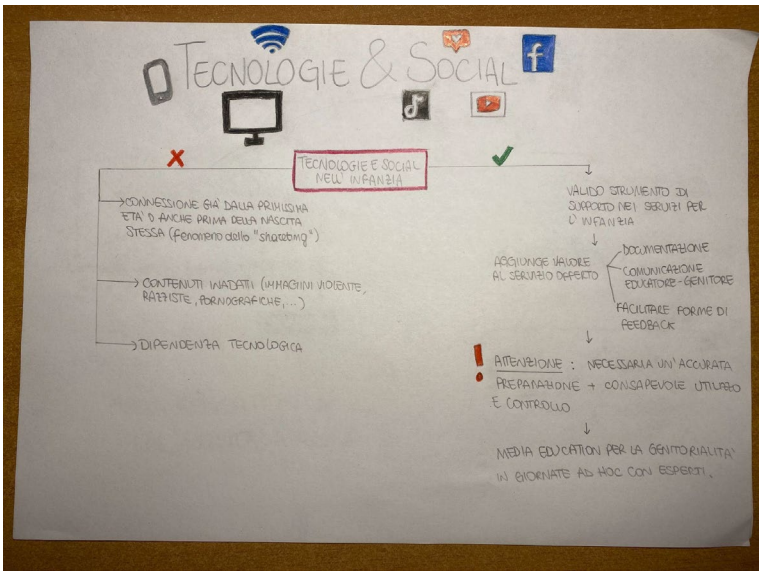
This paper investigated the fundamental aspects of teaching and focused on promoting the refinement of argumentative skills in terms of text comprehension, map construction, final learning, and the general skill of critical thinking.

Starting with the two subsidiary questions, initial analyses suggested that AMs can support the development of argumentative text comprehension and critical thinking in students. An improvement was seen in the ability to correctly name and describe the structural parts of an argumentative text. There was also an improvement in understanding the rules for making a proper AM.

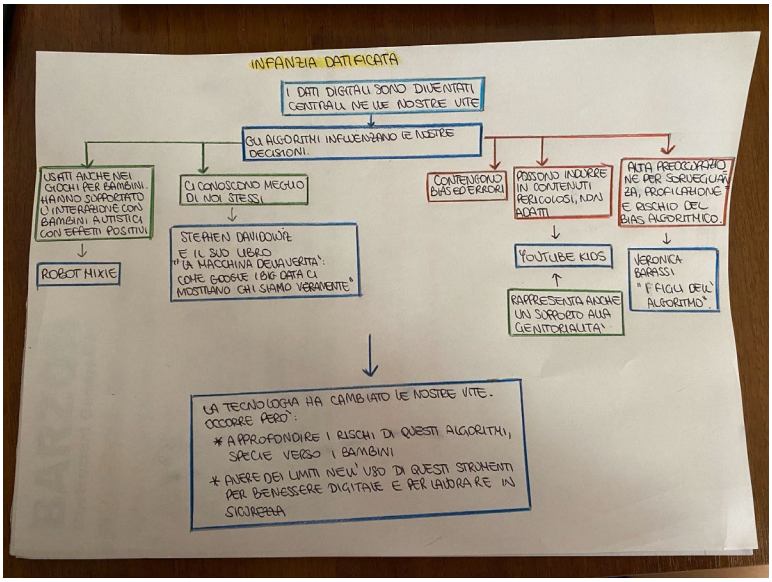
Figure 4 is an illustration of the students' pre- and post-experimentation AMs. The acquisition of this methodology seems to have enabled even the acquisition of a technique for better visualization of the argumentative chain of reasoning.

The images show the rendering of the learned AMs technique. Previously, information was collected in no real order. Subsequently, they were arranged according to the rules of the construction of the AMs. The organization of the main components of the text with colors and precise positions helps with more immediate visualization and reconstruction of the meaning of the text.

The data also certainly showed a growing awareness of the importance of critically analyzing what one reads to form a critical opinion about it. In this case, as mentioned earlier, there was a gradual improvement between pre- and post-intervention time and great significance was also noted. Critical thinking remains one of the most complex and layered of the known skills, and although the results give hope that we are on the right track, there remains a need to continue to explore possible methodologies and uses further towards a well-rounded critical education.



Pre-intervention map



Post-intervention map

Figure 4. Students' AM maps

Turning to the in-depth investigation of the supposed correlation between the two previously discussed skills and the final learning outcomes, the regression model data allowed for the following general picture to be reworked. As a result of the significant correlation with the results inherent in the AMs building skills and the promising data of their relevance in the final performance, a positive pathway inflection with the AMs about the improvement of final learning could be evidenced. Our intervention allowed us to reflect on how the specific use of an instructional approach and tool, which facilitates the reconstruction and visualization of complex thinking, positively influences student learning, not only concurrently with the intervention itself but also with subsequent course learning.

However, it should be noted that the regression model as a whole explained only a small percentage of the variance in the results, which could indicate the presence of other factors influencing students'

final scores that were not accounted for in the model. Hence, it is necessary to reflect on the idea that, although the tool has emerged functional for the stimulation of improved final performance, it is extremely important to deepen, also qualitatively, the most useful and most critical aspects of AMs-based teaching to refine operationalization in research and educational practice.

Another interesting construct that was investigated with a regression model was the level of critical thinking. An attempt was made to investigate and promote the construct of critical thinking as a response variable as one of the cornerstones of active teaching.

From the correlation results first and regression results later, a positive relevance and influence of text comprehension skills on improving the level of critical thinking of students emerged in a hybrid context of learning. Higher levels of proficiency in recognizing the structural elements of an argumentative text were associated with higher levels of improvement in understanding and reconstructing the opinion expressed in the text, as well as in formulating one's own thinking about it. The extreme interrelationship between PostTest_Lab and Post_CriThink traced, however, raised questions about whether the two measurement methods employed to investigate the two variables were perhaps too similar, or whether different tests of critical thinking assessment are needed.

Two other constructs investigated referred to argumentative skills developed in a free, online writing moment (forum) and self-assessment results. Regarding the online forum, it was investigated how much the acquisition of knowledge and skills, such as text comprehension and map construction, affected argumentative skills related to free production on forums.

Hence, it appeared that in a free-writing, online unstructured test, the construction of AMs and the skills related to them, seemed to have had a significant impact, retracing the studies in the field with analog and digital tests. In this regard, we take up previous studies such as those by Dwyer et al. (2010, 2013). Here AMs in the form of map reading had been found to be more functional than text reading in teaching reading comprehension and writing. Positive and interesting results were also reported in a more recent study by Malmir and Khosravi (2018). Indeed, in this study, some argumentative mapping techniques, such as locating relationships, sorting information, and brainstorming, supported the writing process. The AMs provide a coherent and cohesive layout that activates EFL learners' comprehension, language production, and writing process.

Maps involve structuring and analytical skills, which can support reworking during creative, free writing. Our results demonstrate that both analog and digital writing benefit from AM's exercises, based on hybrid environments and instruments supporting the achievement of skills. In contrast, the non-significant correlation with critical thinking outcomes highlighted a possible challenge to critical thinking as a proxy variable for text comprehension and corresponding map construction. This is perhaps due to the use of only one assessment tool, such as the holistically adapted rubric, which, as the only tool used, failed to fully reflect small elements of knowledge acquisition.

Regarding subjective and personal perceptions of improvement, the fact that there was no significant correlation between these results and the actual results of the structured test on knowledge acquisition and map construction showed that this instrument was probably not sensitive enough to show how the experiment affected students.

It seems that when they rate very well, there are quite low scores, and vice versa. Some authors in the literature suggest that this bias effect may be due to "regression towards the mean" (Breidert & Fite, 2009, p. 4). The accuracy of self-assessment seems to moderate with skill level. As a student's skill level increases, the accuracy of self-assessment increases. People with low actual competence assess themselves above average, while people gifted in some skills assess themselves more accurately and lower (Breidert & Fite, 2009; Lew et al., 2010).

Kruger and Dunning (1999, 2002) explained the relationship between skill level (competence), skill esteem (the level of competence), and skill esteem (confidence), pointing out that people who have become more competent are more confident but accurate than those who are less competent and are

overconfident. Our intervention does not seem exempt from this effect. This might also justify the results obtained about the correlation between self-assessment and students' final performance.

Another aspect that may have been influential is related to what our self-assessment instrument sought to investigate – students were asked to comment on the impact of the intervention on their knowledge and the transferability of that knowledge to other areas. The extent of the argumentative skills acquired was not investigated. Thus, a large gap was revealed to be filled, and the component of surprise and discovery of a hitherto unknown element may have taken over. This may have led the lower-performing students to self-evaluate more accurately the idea of having acquired new skills and tools.

Finally, it is relevant to consider some future lines of research born out of the very limitations of our study. First, it might be interesting to integrate qualitative tools to investigate aspects that we may have missed, and that might better explain the need for good final performance. Also, in this way, we could move closer to refining the methodology of AMs, so that they also support long-term learning. In this regard, it might also be interesting to reformulate a self-assessment tool that would dwell more on the argumentative skills acquired during the course. This could prove useful in capturing shades of subjective perceptions, not only about teaching methodology but also about how students themselves learn.

Another necessary step to be put in place is to reconsider the holistic rubric used to assess critical thinking, trying to consider a reconceptualization of it or other tools to go alongside it or replace it, to move towards a more in-depth assessment of this construct.

Building on our findings, practical implications can also be hypothesized for educators who come to think about and design hybrid courses using methodologies such as AMs. These are:

- First, Italian universities could consider the activation and research-based monitoring of pathways for the development of argumentative skills related to critical thinking and education. These can be integrated and enrich the curricula offered to students.
- Second, there is a need to think about pathways specifically for the different disciplinary areas covered. Each discipline has peculiarities, so there are specific educational requirements for the types of argumentative texts and exercises. Hence, the idea of having to rethink future activities, both cross-curricularly and specifically.
- Finally, more connections with the library system could be considered of for the activation of random and longitudinal experimental programs conducted with and by faculty with a focus on the development of information literacy.

Future research can build on precisely these points of criticality and move towards a well-rounded investigation of AMs as a foundational part of didactics aimed not only at learning and developing specific skills, such as argumentative and critical thinking skills, as in this case but also at acquiring transversal skills that can be used in other learning moments as well.

REFERENCES

Alotto, P. (2021). Laboratorio di argomentazione: Guida al critical thinking e all'argument thinking [Argumentation workshop: Guide to critical thinking and argument thinking]. In A. Sani & A. Linguiti (Eds.), *Sinapsi. Storia della filosofia. Protagonisti, percorsi, connessioni.* La Scuola editrice.

ANVUR. (2021a, October 20th). *TECO - TEst sulle COmpetenze: FAQ TECO* [TECO - TEst on COmpetencies: TECO FAQ]. https://www.anvur.it/wp-content/uploads/2021/02/FAQ-TECO-2020_2021.pdf

ANVUR. (2021b, October 20th). *Rilevazione TECO 2020/2021* [TECO 2020/2021 survey]. https://www.anvur.it/wp-content/uploads/2021/12/TECO-In_forma-2020_2021_211216.pdf

Argument Map. (2023, December 13th). *Representing an argument as an argument map.* In Wikipedia. https://en.wikipedia.org/wiki/Argument_map

Breidert, J., & Fite, J. (2009). *Self-assessment: Review and implications for training*. U.S. Army Research Institute for the Behavioral and Social Sciences. https://apps.dtic.mil/sti/pdfs/ADA501346.pdf

Buckingham, D. (2003). *Media Education: Literacy, Learning and Contemporary Culture*. Wiley.

Buckingham, D. (2006). Defining digital literacy – What do young people need to know about digital media? *Nordic Journal of Digital Literacy*, *1*(4), 263–277. https://doi.org/10.18261/ISSN1891-943X-2006-04-03

Buckingham, D. (2009). *The future of media literacy in the digital age: Some challenges for policy and practice*. Medienimpulse.

Buckingham, D. (2015). Do we really need Media Education 2.0? Teaching media in the age of participatory culture. In T. B. Lin, V. Chen, & C. S. Chai (Eds.). *New media and learning in the 21st Century. A socio-cultural perspective* (pp. 9–21). Springer. https://doi.org/10.1007/978-981-287-326-2_2

Canale, D., Frigerio, A., Giovanni, T., & Roberto, C. (2021). *Critical thinking: An introduction*. Egea.

Carrington, M., Chen, R., Davies, M., Kaur, J., & Neville, B. (2011). The effectiveness of a single intervention of computer-aided argument mapping in a marketing and a financial accounting subject. *Higher Education Research & Development*, *30*, 387–403. https://doi.org/10.1080/07294360.2011.559197

Chounta, I., McLaren, B. M., & Harrell, M. (2017). Building arguments together or alone? Using learning analytics to study the collaborative construction of argument diagrams. In B. K. Smith, M. Borge, E. Mercier, & K. Y. Lim (Eds.), *Proceedings of the 12th International Conference on Computer Supported Collaborative Learning* (pp. 589–592). International Society of the Learning Sciences.

Coates, H., Xie, Z., & Hong, X. (2021). Engaging transformed fundamentals to design global hybrid higher education. *Studies in Higher Education*, *46*(1), 166–176. https://doi.org/10.1080/03075079.2020.1859683

Colombo, A. (2018). Il testo argomentativo: Presupposti pedagogici e modelli di analisi [The argumentative text: Pedagogical assumptions and models of analysis]. In A. Colombo (Ed.), *Quaderni del Giscel: Vol. 11. I pro e i contro* (pp. 59–84). La Nuova Italia. https://giscel.it/wp-content/uploads/2018/04/Adriano-Colombo-Il-testo-argomentativo-Presupposti-pedagogici-e-modelli-di-analisi.pdf

Commissione Europea. (2010). *Europa 2020: Una strategia per una crescita intelligente, sostenibile e inclusiva* [Europe 2020: A strategy for smart, sustainable and inclusive growth]. EUR-Lex. https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:2020:FIN:it:PDF

Crudele, F., & Raffaghelli, J. E. (2023). *Argument Maps as tools to support the development of new media literacies: A systematic review* [data set]. Zenodo. https://doi.org/10.5281/zenodo.7729327

Crudele, F., & Raffaghelli, J. E. (in press). *Argument maps as tools to support the development of new media literacies: A systematic review*. Formazione & Insegnamento.

Cumming, G. (2014). The new statistics: why and how. *Psychological Science*, 25(1), 7-29. https://doi.org/10.1177/0956797613504966

Davies, M. (2011). Concept mapping, mind mapping and argument mapping: What are the differences and do they matter? *Higher Education*, *62*, 279–301. https://doi.org/10.1007/s10734-010-9387-6

Davies, M. (2012). Computer-aided argument mapping and the teaching of critical thinking: Part II. *Inquiry: Critical Thinking across the Disciplines*, *27*(3), 16–28. https://doi.org/10.5840/inquiryct201227317

Dwyer, C. P., Hogan, M. J., & Stewart, I. (2010). The evaluation of argument mapping as a learning tool: Comparing the effects of map reading versus text reading on comprehension and recall of arguments. *Thinking Skills and Creativity*, *5*(1), 16–22. https://doi.org/10.1016/j.tsc.2009.05.001

Dwyer, C. P., Hogan, M. J., & Stewart, I. (2013). An examination of the effects of argument mapping on students' memory and comprehension performance. *Thinking Skills and Creativity*, 8, 11–24. https://doi.org/10.1016/j.tsc.2012.12.002

Eftekhari, M., & Sotoudehnama, E. (2018). Effectiveness of computer-assisted argument mapping for comprehension, recall, and retention. *ReCALL, 30*(3), 337–354. https://doi.org/10.1017/S0958344017000337

Facione, P. A., & Facione, N. C. (2014). *The Holistic Critical Thinking Scoring Rubric – HCTSR*. The California Academic Press/Insight Assessment.

Fan, C.-Y., & Chen, G.-D. (2021). A scaffolding tool to assist learners in argumentative writing. *Computer Assisted Language Learning*, *34*(1–2), 159–183. https://doi.org/10.1080/09588221.2019.1660685

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Sage.

Ganino, G. (2020). Riduzione carico cognitivo estraneo e apprendimenti multimediali in un ambiente di web conference [Reducing extraneous cognitive load and multimedia learning in a web conference environment]. *Reports on E-Learning, Media and Education Meetings*, *8*(1), 234–239. https://www.je-lks.org/ojs/index.php/R-EMEM/article/view/1135272

Harrell, M. (2022). Representing the structure of a debate. *Argumentation*, *36*, 595–610. https://doi.org/10.1007/s10503-022-09586-2

Harrell, M., & Wetzel, D. (2015). Using argument diagramming to teach critical thinking in a first-year writing course. In M. Davies, & R. Barnett (Eds.), *The Palgrave handbook of critical thinking in higher education* (pp. 213–232). Palgrave Macmillan. https://doi.org/10.1057/9781137378057_14

Hobbs, R. (2020). Propaganda in an age of algorithmic personalization: Expanding literacy research and practice. *Reading Research Quarterly*, *55*(3), 521–533. https://doi.org/10.1002/rrq.301

Hobbs, R., Kanižaj, I., & Pereira, L. (2019). Digital literacy and propaganda. *Medijske Studije*, *10*(19), 1–7. https://doi.org/10.20901/ms

Hoffman, J. I. E. (2019). Chapter 26 - Analysis of Variance. II. More Complex Forms. In J. I. E. Hoffman (Ed.), *Basic Biostatistics for Medical and Biomedical Practitioners (Second Edition)*. Academic Press. https://doi.org/10.1016/B978-0-12-817084-7.00026-7

INVALSI. (2015). Rilevazioni nazionali degli apprendimenti 2015-2016: Rapporto risultati [National Learning Surveys 2015-2016: Results report]. https://www.invalsi.it/invalsi/doc_evidenza/2016/07_Rapporto_Prove_INVALSI_2016.pdf

INVALSI. (2018). *Rapporto nazionale: Rapporto prove INVALSI 2018* [National report: INVALSI Testing Report 2018]. https://www.invalsi.it/invalsi/doc_evidenza/2018/Rapporto_prove_INVALSI_2018.pdf

Iordanou, K., & Rapanta, C. (2021). "Argue with me": A method for developing argument skills. *Frontiers in Psychology, 12*. https://doi.org/10.3389/fpsyg.2021.631203

Iswati, H. D., & Purwati, O. (2022). Improving students' critical thinking through guided discovery learning method in argumentative texts reading. *ELS Journal on Interdisciplinary Studies in Humanities*, *5*(4), 701–705. https://doi.org/10.34050/elsjish.v5i4.24826

Kabataş MemiŞ, E., & Karakuş, E. (2021). An evaluation of academic achievements through the use of argument and concept maps embedded in argumentation based inquiry. *Asia Pacific Education Review, 22*(3), 463–481. https://doi.org/10.1007/s12564-021-09679-9

Kaeppel, K. (2021). The influence of collaborative argument mapping on college students' critical thinking about contentious arguments. *Thinking Skills and Creativity, 40*, 100809. https://doi.org/10.1016/j.tsc.2021.100809

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*(6), 1121–1134. https://doi.org/10.1037//0022-3514.77.6.1121

Kruger, J., & Dunning, D. (2002). Unskilled and unaware - But why? A reply to Krueger and Mueller (2002). *Journal of Personality and Social Psychology*, *82*, 189–192. https://doi.org/10.1037//0022-3514.82.2.189

Kuhn, D. (2010). Thinking as argument. *Harvard Educational Review*, *62*(2), 155–179. https://doi.org/10.17763/haer.62.2.9r424r0113t670l1

Kuhn, D. (2018). A role for reasoning in a dialogic approach to critical thinking. *Topoi, 37*(1), 121–128. https://doi.org/10.1007/s11245-016-9373-4

Lew, M. D. N., Alwis, W. A. M., & Schmidt, H. G. (2010). Accuracy of students' self-assessment and their beliefs about its utility. *Assessment & Evaluation in Higher Education*, *35*, 135–156. https://doi.org/10.1080/02602930802687737

Lidåker, T. (2018). *The potential of argument mapping as a tool for teaching critical thinking in secondary school.* [Master dissertation, Linkoping University]. http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-149438

Lo Feudo, G. (2018). Serialità narrativa televisiva: Linguaggi e testualità [Television narrative seriality: Languages and textuality]. *Filosofi(e) Semiotiche*, *5*(2), 6. https://doi.org/20.500.11770/288939

Macgilchrist, F. (2021). Theories of postdigital heterogeneity: Implications for research on education and datafication. *Postdigital Science and Education*, *3*(3), 660–667. https://doi.org/10.1007/s42438-021-00232-w

Malmir, A., & Khosravi, F. (2018). The effect of argument mapping instruction on L2 writing achievement across writing tasks and writing components: A case of Iranian EFL learners. *Applied Research on English Language*, *7*(4), 515–540. https://doi.org/10.22108/are.2018.111870.1318

Mochizuki, T., Nishimori, T., Tsubakimoto, M., Oura, H., Sato, T., Johansson, H., Nakahara, J., & Yamauchi, Y. (2019). Development of software to support argumentative reading and writing by means of creating a graphic organizer from an electronic text. *Educational Technology Research and Development*, *67*, 1197–1230 https://doi.org/10.1007/s11423-019-09676-1

Moretti, G. (2010). Lucia Lumbelli (2009). La comprensione come problema. Il punto di vista cognitive [Lucia Lumbelli (2009). Understanding as a problem. The cognitive point of view]. *Journal of Educational, Cultural and Psychological Studies*, *1*(2), 201–207. https://www.ledonline.it/index.php/ECPS-Journal/article/view/121

OCSE-PISA. (2015). *I risultati OCSE PISA 2015: Descrizione dei livelli di competenza nella scala di literacy di Lettura* [OECD PISA 2015 results: Description of proficiency levels on the Reading literacy scale]. INVALSI. https://www.invalsi.it/invalsi/ri/pisa2015/doc/rapporto_PISA_2015.pdf

OECD. (2020). *Education at a glance 2020: OECD indicators.* https://doi.org/10.1787/69096873-en

OECD. (2021). *Education at a glance 2021: OECD indicators.* https://doi.org/10.1787/b35a14e5-en

Özdemir, S. (2018). The effect of argumentative text pattern teaching on success of constituting argumentative text elements. *World Journal of Education*, 8(5), 112–122. https://doi.org/10.5430/wje.v8n5p112

Simari, G. R., & Rahwan, I. (2009). *Argumentation in artificial intelligence.* Springer. https://doi.org/10.1007/978-0-387-98197-0

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1016/0364-0213(88)90023-7

Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 19–30). Cambridge University Press. https://doi.org/10.1017/CBO9780511816819.003

van der Brugge, E. (2018). *The use of argument mapping in improving critical thinking* [Doctoral dissertation, University of Melbourne]. http://hdl.handle.net/11343/214519

Wambsganss, T., Niklaus, C., Cetto, M., Söllner, M., Handschuh, S., & Leimeister, J. M. (2020). AL: An adaptive learning support system for argumentation skills. In R. Bernhaupt, F. F. Mueller, D. Verweij, J. Andres, J. McGrenere, A. Cockburn, I. Avellino, A. Goguey, P. Bjørn, S. Zhao, B. P. Samson & R. Kocielnik (Eds.), *Proceedings of the CHI Conference on Human Factors in Computing Systems* (pp. 1–14). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376732

APPENDIX A

THE SEMI-STRUCTURED SURVEY INSTRUMENT WITH CLOSED STIMULUS QUESTIONS

Course of Study in Education and Training Sciences (EPI) **Teaching of Planning, Documentation and Evaluation For Early Childhood** **AA 2021/2022**	
Name and Surname	
Read carefully the argumentative text you received for the day's workshop activity. Then complete the following questions, and then upload the file to the related task space.	
1st question	What is the topic (**problem**) discussed in the text?
2nd question	What is the author's opinion (**thesis**)?
3rd question	Identify the **arguments that seem to support** the author's opinion and quote the sentences from the text below.
4th question	Identify, if any, the **opinion contrary** to the author's opinion (**antithesis**) and report the sentences from the text below.
5th question	Identify the evidence (**foundation**) of the argument.
6th question	What, in your opinion, is the **conclusion** reached by the author?
7th question	What is your **opinion** about it? And how much has it changed as a result of reading the author's point of view?

APPENDIX B

CHECK-LIST FOR IDENTIFYING THE STRUCTURAL ELEMENTS OF AN ARGUMENTATIVE TEXT

Questions	**Component identified**	**Component partially identified**	**Component not identified**
	1 point	**0.5 points**	**0 points**
What is the problem discussed in the text?			
What is the author's opinion, the thesis?			
What arguments seem to you to support the author's opinion?			
What, if any, is the contrary opinion to the author's?			
What is the basis of the argument?			
What is the conclusion reached by the author?			
What is your opinion on the matter? And how much has it changed as a result of reading the author's point of view?			

APPENDIX C

AN ADAPTED VERSION OF THE HOLISTIC CRITICAL THINKING SCORING RUBRIC (HCTSR)

	Strong	**Acceptable**	**Weak**	**Absent**
Identify important information (Questions 1-2)	Carefully interprets evidence, statements, questions, and information.	Interprets evidence, statements, questions, and information accurately.	Partially interprets evidence, questions, information, or the other person's point of view.	Offers a biased interpretation of evidence, questions, information, or the other person's point of view.
	2.5 points	**1.75 points**	**1.25 points**	**0.5 points**
Identify arguments and counter-arguments or alternative points of view (Questions 3-5)	Identifies the most important arguments for or against, thoroughly analyzing the most significant alternative viewpoints.	Identifies only strong and obvious counter-arguments and ignores or superficially evaluates alternative viewpoints.	Does not hastily identify or set aside strong and permanent counter-arguments.	Does not hastily identify or set aside strong and permanent counter-arguments. Ignores or superficially evaluates alternative viewpoints.
	2.5 points	**1.75 points**	**1.25 points**	**0.5 points**
Draw conclusions and explain reasons (Question 6)	Draws conclusions that are justified, sensible, and not fallacious. Justifies key findings and procedures, explaining assumptions and reasons.	Draws conclusions that are justified and not fallacious. Motivates few results or procedures, rarely explains reasons.	Draws unwarranted or fallacious conclusions. Justifies a few results or procedures and rarely explains reasons.	Argues using fallacious reasons, does not justify results or procedures, and does not explain the reasons.
	2.5 points	**1.75 points**	**1.25 points**	**0.5 points**
Understand and modify one's opinion based on evidence (Question 7)	Fairly follows where evidence and reasons lead him.	It follows fairly where evidence and reasons lead.	Regardless of evidence and reasons, maintains or defends viewpoints based on personal interests or preconceptions.	Regardless of evidence and reasons, maintains or defends viewpoints based on personal interests or preconceptions. Shows closed-mindedness and hostility toward reason.
	2.5 points	**1.75 points**	**1.25 points**	**0.5 points**

APPENDIX *D*

A CLOSER LOOK AT THE STATISTICAL RESULTS

This appendix contains copies of earlier tables with a more detailed discussion of each.

(Copy of Table 3) Descriptive statistics - text comprehension test

Situation	Mean	Std. Dev	Min	Q1	Median	Q3	Max	Skewness	Kurtosis
Students' comprehension of the text									
S1	3.95	1.43	1.00	2.50	4.00	5.00	7.00	-0.04	-0.67
S2	4.23	1.35	1.00	3.25	4.50	5.50	7.00	-0.33	-0.43
S3	4.48	1.34	1.00	3.50	4.50	5.50	7.00	-0.34	-0.46

In this regard, the one-way repeated-measures ANOVA test was considered for the analyses. However, there were violations of the conditions for conducting the test, as can be verified from the analyses. Therefore, Friedman's equivalent nonparametric test (Hoffman, 2019) was applied and a moderately significant difference in text comprehension was found (X2 (2), N=100) = 12.101, p<.01). A subsequent post-hoc test (non-parametric Wilcoxon), revealed a significant difference between pre-test and post-test scores (p<.001). This allowed us to read an improvement between the two moments and note the significance of the effect of AMs in facilitating the identification of the structural elements of the argumentative text and subsequent comprehension of the text.

Ultimately, by estimating and applying the effect size, it was possible to confirm a moderate effect between the pre- and post-test ($\eta^2 = 0.397$). These deeper analyses, therefore, allowed a glimpse of how AMs supported better and easier identification of the structural components of an argumentative text.

(Copy of Table 4) Descriptive statistics - argumentative map construction

Situation	Mean	Std. Dev.	Min	Q1	Median	Q3	Max	Skewness	Kurtosis
Completeness of maps constructed by students									
M1	4.46	1.10	2.50	3.50	4.00	5.50	7.00	0.46	-0.97
M2	4.58	0.95	3.00	4.00	4.50	5.50	6.50	0.21	-0.96
M3	4.68	1.22	2.00	4.00	4.50	6.00	7.00	0.09	-0.91

In this case, widespread low significance was found (X2 (2), N=100) = 3.4463, p>.01). The subsequent post-hoc test found no significant difference between testing moments (p>.05). Then applying the effect size, a small effect size was found and confirmed in all three moments (respectively, η^2 1 = 0.190; η^2 2 = 0.179; η^2 3 = 0.008).

(Copy of Table 5) Descriptive statistics - level of critical thinking

	Mean	Std. Dev	Min	Q1	Median	Q3	Max	Skewness	Kurtosis
Level of Critical Thinking with HCTSR Rubric									
CT1	6.66	1.44	4.00	5.50	6.50	7.75	10.00	0.28	-0.62
CT2	7.78	1.07	4.25	7.25	7.75	8.50	10.00	-0.33	0.45
CT3	8.00	1.03	5.50	7.25	8.00	8.50	10.00	0.02	-0.51

The positive results obtained from the descriptive analyses prompted an investigation of the magnitude of the effect obtained.

One-way ANOVA analysis here revealed a significant difference in the critical thinking data ($F(2)=36.85$, $p<.001$). The subsequent post-hoc test then found this significant difference at two testing times – between CT1-CT2 ($p <.001$) and between CT1-CT3 ($p<.001$). This could indicate an increasing improvement in understanding the opinion expressed in an argumentative text and in formulating one's own thinking about it. Here, with effect size, a large effect was found between these two testing moments ($r1 = 0.618$; $r2 = 0.709$, respectively).

Preparatory phase: Spearman Correlation

The results showed no significant Spearman correlation between Forum_Post, the final detection of argumentative skills in the forum, and Post_CriThink, the final level of critical thinking ($rho = 0.12$, $p = 0.27$). For the rest of the variables, a positive significant correlation was observed with Map3b_Lab, the final skill of constructing AMs ($rho = 0.36$, $p<0.001$), and a positive moderating correlation with PostTest_Lab, the final moment of text comprehension ($rho = 0.25$, $p<0.01$).

The second aspect investigated was students' subjective perception, or self-assessment. Spearman's rank correlation showed a negative correlation between Autoval_Post and PostTest_Lab ($rho = -0.07$, $p = 0.59$) and Post_CriThink ($rho = -0.04$, $p = 0.75$). For the rest of the variables, a positive but non-significant correlation emerges (Map3b_Lab, $rho = 0.003$, $p = 0.98$; Grade_Fin, $rho = 0.04$, $p = 0.76$).

Regarding final performance, Spearman correlation showed a significant positive correlation with Map3b_Lab ($rho = 0.37$, $p<0.001$) and a moderate positive correlation with the variables PostTest_Lab ($rho = 0.26$, $p<0.05$) and Post_CriThink ($rho = 0.37$, $p<0.05$). A heatmap graph was used at this point, representing the correlation matrix more clearly and intuitively. This revealed the strong positive correlation between PreTest_Lab and Pre_Crithink ($rho = 0.78$), and PostTest_Lab and Post_Crithink ($rho = 0.88$). These results might suggest excessive linearity between the two predictors, perhaps because of the strong connection and interconnectedness of the data collection and analysis tools. Hence, future investigations must keep in mind that the two predictors may be interchangeable.

Regression phase: regression models

Exploratory analysis to test the model's hypotheses did not reveal any violations. In the set of explanatory variables, the Map3b_Lab predictor contributed significantly to the model ($\beta = 1.032$, $t = 2.796$, $p<0.01$). None of the other variables reached statistical significance. The overall model showed a good fit to the data (F-statistic: 3.648 on 4 and 77 DF, p-value <0.01, $AR2 = 0.12$). However, it is important to keep in mind that the four predictor variables in the model can only account for 12% of the variance in students' final scores. From these assumptions, the pre-test moments were removed from the model as irrelevant and incident predictors.

The exploratory analysis for testing the hypotheses of the newly formulated model version did not violate any conditions. In the set of explanatory variables, the t-statistic returned the significance of the explanatory variable Map3b as a predictor variable of the model ($\beta = 1.010$, $t = 2.948$, $p<0.01$). However, the model would appear to explain only 14% of the variance in the results. There could, therefore, be other variables or factors that influenced the response and were not considered in the model.

The second regression model (open data on Zenodo, anonymized) was constructed, including PreTest_Lab, PostTest_Lab, Map3b_Lab and Forum_Post as explanatory variables, and the Post_Crithink as a response variable. Condition testing analysis found a violation of the normal distribution of residuals and two influential observations that could significantly affect the regression model. Bootstrap confidence intervals for these predictors were, therefore, identified. Comparing these intervals with the confidence interval found using the plug-in approach, it was found that all bootstrap confidence intervals are very close to the plug-in confidence intervals, suggesting that there is no non-normal distribution problem in the model.

Overall, the regression model showed adequate prediction ability of the dependent variable (F-statistic: 30.01 on 4 and 83 DF, $p<0.001$, $AR^2 = 0.57$). In the set of explanatory variables, the PostTest_Lab predictor appears to contribute significantly to the model ($\beta = 0.54$, $p<0.001$). The other variables do not appear to be significantly associated with the response variable. Among others, the predictor Forum_Post was found to be negatively associated with Post_CriThink ($\beta = -0.09$, $p = 0.20$).

The model was then adjusted, removing Forum_Post as a predictor and investigating the influence of the explanatory variables PreTest_Lab, PostTest_Lab, Map1b_Lab and Map3b_Lab. Analysis for condition testing of the new model version shows a violation of the condition of homoscedasticity and normal distribution of residuals.

Having identified the bootstrap confidence intervals, it was found that the bootstrap results are more or less all close except for the pre-test moments: PreTest_Lab (plug-in confidence interval -0.004 to 0.203 and bootstrap confidence interval -0.265 to 0.0001) and Map1b_Lab (plug-in confidence interval -0.132 to 0.067 and bootstrap confidence interval -0.290 to -0.0006). This suggests a normal distribution problem in the model.

The regression equation was significant (F: 39.45 on 4 and 94 DF, $p<0.001$, AR2 = 0.61) explaining about 61% of the variance in Post_CriThink scores. The PostTest_Lab predictor contributed significantly to the model ($\beta = 0.536$, $t(94) = 9.484$, $p<0.001$). The other explanatory variables do not appear to have contributed significantly to the model.

APPENDIX E:

SPEARMAN'S CORRELATION: APA STYLE TABLE

Means, standard deviations, and correlations with confidence intervals

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10
1. PreTest_Lab	3.95	1.43										
2. Map1b_Lab	4.46	1.10	.27** [.08, .44]									
3. PostTest_Lab	4.48	1.34	.44** [.27, .59]	.30** [.11, .47]								
4. Map3b_Lab	4.67	1.22	.29** [.10, .46]	.37** [.19, .53]	.39** [.21, .54]							
5. Pre_CriThink	6.66	1.44	.88** [.82, .91]	.24* [.05, .42]	.32** [.13, .49]	.37** [.21, .50]						
6. Post_CriThink	8.00	1.03	.44** [.27, .59]	.15 [-.05, .34]	.77** [.67, .84]	.39** [.21, .55]	.43** [.26, .58]					
7. Forum_Pre	7.29	1.43	-.01 [-.22, .20]	.02 [-.20, .23]	.22* [.01, .42]	.20 02, .39]	.05 16, .26]	.28* [.06, .46]				
8. Forum_Post	8.05	1.21	.13 [-.08, .33]	.26* [.06, .44]	.25* [.06, .43]	.36** [.19, .52]	.16 [-.05, .35]	.12 06, .29]	.21 00, .41]			

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10
9. Autoval_Pre	2.58	0.60	-.06 [-.29, .17]	-.07 [-.29, .17]	-.24* [-.45, -.01]	-.05 28, .19]	-.09 31, .15]	-.19 40, .05]	-.12 34, .12]	.03 [-.20, .26]		
10.Autoval_Post	4.00	0.39	-.05 [-.28, .18]	-.09 [-.31, .15]	-.07 [-.26, .13]	.00 21, .23]	-.06 29, .17]	-.04 24, .16]	.15 08, .37]	-.03 [-.26, .20]	.27* [.05, .47]	
11. Grade_Fin	20.81	3.63	.16 [-.05, .37]	.16 [-.06, .36]	.26* [.07, .42]	.37** [.19, .53]	.17 05, .37]	.24* [.05, .40]	.27* [.05, .46]	.18 [-.04, .38]	-.00 24, .23]	.04 [-.15, .23]

Note. *M* and *SD* are used to represent mean and standard deviation, respectively. Values in square brackets indicate the 95% confidence interval for each correlation. The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming, 2014).
* indicates $p < .05$. ** indicates $p < .01$

AUTHORS

Francesca Crudele is a third-year PhD student in Pedagogical Science in the FISPPA Department at the University of Padua. Her doctoral research is about the development of a new idea of argument maps to improve undergraduates' argumentative skills and critical thinking in new media contexts.

Juliana Elisa Raffaghelli is an assistant professor in educational research and experimental pedagogy at the University of Padua, Italy. Her research has focused on educators' critical digital and data literacy. She actively practices the values of open science and education, exploring professional identities and agency in a context of private platform dominance and datafication.