Form@re
Open Journal per la formazione in rete

Articles

New job skills: critical thinking analysis in higher education and engineering

Competenze per nuovi lavori: analisi del pensiero critico nell'educazione superiore e ingegneria

Antonella Poce[a], Carlo De Medio[b,1]

[a] *University of Rome Tor Vergata, antonella.poce@uniroma2.it*
[b] *University of Rome Tor Vergata, carlo.demedio.572666@uniroma2.eu*

Abstract

This article presents an approach for the assessment of critical thinking based on the analysis of open-ended answers, highlighting its advantages over closed-ended tests. The need to develop and measure reasoning skills also in STEM students is highlighted. During a workshop addressed to engineering students, open-ended questions were administered to assess critical thinking both through a manual scoring rubric system and through an automated prototype of Natural Language Processing (NLP). This approach answers two main research questions: the degree of concordance between manual and automatic assessment; the effectiveness of the workshop in enhancing critical thinking in a STEM context. An encouraging alignment between the two forms of assessment is found and how targeted training-evaluation activities can foster the acquisition of critical skills also in subject specific field is underlined.

Keywords: critical thinking; Natural Language Processing; AI; soft skills.

Sintesi

L'articolo presenta un approccio per la valutazione del pensiero critico basato sull'analisi delle risposte aperte, evidenziandone i vantaggi rispetto ai test a risposta chiusa. Si sottolinea la necessità di sviluppare e misurare competenze di ragionamento anche negli studenti delle STEM. Durante un workshop rivolto a studenti di ingegneria, sono state somministrate domande a risposta aperta per valutare il pensiero critico sia attraverso un sistema di rubriche a punteggio manuale sia attraverso un prototipo automatizzato di elaborazione del linguaggio naturale (NLP). Questo approccio risponde a due principali domande di ricerca: il grado di concordanza tra la valutazione manuale e quella automatica; l'efficacia del workshop nel migliorare il pensiero critico in un contesto STEM. Si riscontra un allineamento incoraggiante tra le due forme di valutazione e si sottolinea come attività di formazione-valutazione mirate possano favorire l'acquisizione di competenze critiche anche in ambiti specifici.

Parole chiave: pensiero critico; elaborazione del linguaggio naturale; IA; soft skills.

[1] Antonella Poce chaired the research group that managed the project and is the author of paragraphs 1, 2 and 6; Carlo De Medio authored paragraphs 3, 4 and 5.

1. Introduction

In recent years, the topic of soft skills has become central to academic, political and professional debate (Robles, 2012). With the advent of digitalization, rapid technological changes and the emergence of new forms of work organization, companies require not only hard skills (i.e. technical and specific) but also soft skills, i.e. transversal and relational skills (Heckman & Kautz, 2012). These include communication skills, problem solving, critical thinking, leadership, teamwork, time management and complex situations. According to the *Future of Jobs Report* of the World Economic Forum [WEF] (2020), significant changes in the skills required by the labor market are expected in the coming years, with an increasing emphasis on adaptability, continuous learning and empathic communication skills. At the same time, various national (Unioncamere & Anpal, 2024) and international (Organization for Economic Co-operation and Development [OECD], 2019) institutions and researches underline the importance of these skills in the most dynamic employment contexts. In Italy, the National Statistics Institute [Istat] (2022) shows how companies in rapidly expanding sectors (technology, advanced manufacturing, digital services, green economy) are particularly interested in integrating soft skills, considered essential to manage complexity and innovation. At a global level, this trend is confirmed by LinkedIn Learning's analyses (2023), according to which employers around the world show increasing attention to training and developing their employees' transversal skills, believing that these skills can guarantee greater added value and improved company competitiveness. The term "soft skills" first appeared in American military training studies in the 1970s, where it distinguished relational skills from technical skills (Klaus, 2010). Today, soft skills are referred to as a set of skills related to emotional intelligence, communication skills and interpersonal relationship management (Goleman, 1995). Compared to hard skills – which refer to technical, measurable and certifiable knowledge and skills – soft skills concern personal, attitudinal and behavioural attributes, which are often more difficult to quantitatively evaluate (Zehr, 1998). Among the various taxonomies and theoretical models, the classification proposed by Heckman and Kautz (2012) stands out, according to which non-cognitive skills – of which soft skills are a part – have a significant impact on a career, productivity levels and quality of life. This vision is consistent with studies on human capital (Becker, 1975) and psychological capital (Luthans et al., 2007), which consider personal skills and characteristics as a real long-term investment in terms of professional success and individual well-being. In a context increasingly characterized by complexity and the need for continuous professional updating, the human capital approach cannot ignore the role of soft skills. They act as a "bridge" between specific (hard) skills and the ability to transfer and apply them in a creative, empathetic and results-oriented way (Trilling & Fadel, 2009). In the context of Industry 4.0 and digital transformation, more and more routine tasks are being automated or delegated to Artificial Intelligence (AI), while workers are required to have a greater ability to solve complex problems, interact with multidisciplinary teams and operate in situations of uncertainty (WEF, 2020). It is in this perspective that soft skills become crucial not only to obtain a first job, but also to maintain one's employability in the long term (Schulz, 2008). The OECD *Skills Outlook* (2019) highlights how the lack of transversal skills represents a significant gap between recent graduates and market needs. This discontinuity is particularly felt in Italy, where many companies complain about the difficulty in finding professionals who, in addition to specialist skills, also possess communication skills, adaptability and critical thinking (Unioncamere & Anpal, 2024). A concrete example is provided by the technology sector: high-tech companies do seek programming and data analysis skills, but often state that the ability to work in a team and flexibility in learning new computer languages are crucial to the success of candidates (Capelli, 2015). Similarly, in the financial sector, recruiters emphasize the importance of relational, negotiation and leadership skills (Roselina, 2009). Eurostat data (2025) show that employment in Europe is growing in sectors with a high content of knowledge and innovation, particularly in professional services, the green economy and the ICT (Information and Communication Technology) sector.

In Italy, Unioncamere and Anpal (2024) estimate that in the next five years the demand for professionals linked to digital transformation, energy transition and innovative services will grow, with a need for transversal skills equal to over 70% of new hires. At the same time, the WEF (2020) predicts that many "traditional" professions will undergo profound transformations, requiring workers to update their skill sets. In the face of this evolution, soft skills are configured as a key factor of professional resilience and the ability to adapt to change, both in terms of internal mobility within organizations and mobility between different sectors (Deming, 2017). According to LinkedIn's Workplace Learning Report (2023), the soft skills most sought after by companies globally include: effective communication**,** teamwork and collaboration**.** time management and organization**,** critical thinking and problem solving**,** leadership and conflict management. The transition from higher education to the workplace is a crucial time in which new graduates are faced with increasingly complex requirements (Barton et al., 2013). Numerous studies highlight how companies, in addition to solid disciplinary skills, prefer candidates who are able to demonstrate proactivity, initiative, and the ability to communicate and collaborate (Robles, 2012). In particular:

- ability to learn quickly: in an ever-changing environment, learning to learn has become an essential cross-curricular skill (Cobo, 2013);
- flexibility and adaptability: ability to reinvent oneself and manage unexpected situations;
- transversal digital skills: although not strictly speaking "soft skills", they are often considered part of the basic skills required in almost all sectors (Van Laar et al., 2017);
- ability to team working: knowing how to work in intercultural and multidisciplinary teams, even remotely (Gilbert, 2012);
- intercultural sensitivity: with the globalization of markets, the propensity to understand and respect different cultures is highly appreciated (Earley & Ang, 2003).

McKinsey's *Skills Gap Report* (Bughin et al., 2018) highlights how a large part of European and American companies have difficulty finding candidates who possess an adequate combination of hard and soft skills. In particular, recruiters highlight how the academic background is often of a good level, but lacks a practical approach, problem solving and relational skills necessary to work effectively within groups (Grugulis & Vincent, 2009).

At an international level, there are significant differences in the emphasis placed on soft skills. In the United States, for example, the university system and the business ecosystem actively collaborate through internships, placements and joint research projects, encouraging the acquisition of relational and leadership skills already during the course of studies (Bettiol & Micelli, 2020). In Asia, countries such as Singapore, South Korea and Japan have launched national "lifelong learning" programmes that aim to strengthen workers' skills to adapt to rapid technological changes (OECD, 2019). This has led to a greater focus on communication skills, problem solving and lifelong learning as a key ingredient to compete on a global scale (Lam, 2019). In Europe, the European Union policies (European Commission [EC], 2020) encourage schools and universities to integrate teaching methods oriented to the development of soft skills, such as group work, project-based learning and business simulation activities (Trilling & Fadel, 2009). However, the academic tradition of many European countries – including Italy – remains partly tied to more transmissive and frontal teaching methodologies, with fewer opportunities to experiment and cultivate transversal skills (Weber et al., 2009). A further comparison can be made between Northen European countries (Sweden, Denmark, Finland) and Southern Europe (Italy, Spain, Greece). In the former, the educational model tends to place greater value on active participation, student autonomy and collaboration, encouraging the development of transversal skills from the early years of education (Blömeke et al., 2017). In the latter, the transition to more interactive and collaborative teaching methodologies proceeds gradually and unevenly. Soft skills play a fundamental role in defining the professional profile of 21st century workers. The evolution of the labor market, characterized by rapid technological and organizational changes, makes a balanced combination of technical skills and transversal skills increasingly necessary. Sector studies, statistical analyses and empirical research converge in underlining how the mastery of relational, communication, leadership and problem solving skills is decisive for entering and maintaining employment. At the international level, although differences emerge in terms of training systems and labor policies, there is substantial agreement on the need to encourage the development of soft skills, both through school and university education and through lifelong learning paths. In the Italian context, the need to reduce the skills mismatch between training supply and business demand represents a priority for both educational institutions and public policies, with the aim of promoting the economic competitiveness and employability of young graduates. From what has emerged, it is clear that there is an opportunity to promote further studies and interventions to define more effective teaching strategies, learning methodologies and soft skills assessment tools. In this way, it will be possible to facilitate the transition of new graduates to the world of work and guarantee a flexible, innovative workforce capable of successfully facing the challenges of the future.

2. Methodology

Critical Thinking (CT) has long been recognized as an essential skill in many professional fields, but its teaching and assessment still pose significant challenges. While many standardized tools (e.g., multiple-choice tests and self-report questionnaires) allow for rapid and large-scale assessments, it is difficult to adequately measure the processes of analysis, argumentation, and reflection that characterize true critical thinking. In this sense, the use of open-ended questions offers the possibility of capturing nuances of students' reasoning that would otherwise be lost, thus allowing for the measurement and promotion of skills of analysis, evaluation, inference, and meta-reflection (Facione, 1990; Poce, 2017).

In this article, we will first explore the importance of using open-ended questions to assess Critical Thinking, referring to the research and methodological approaches described in Poce et al. (2019b). Second, we will discuss the need to foster the development of such skills also in students coming from the so-called "hard sciences" (e.g., engineering, physics, mathematics), where it is often wrongly believed that the mastery of solid hard skills is in itself sufficient to guarantee professional success. Finally, we will describe a workshop addressed to engineering students focused on the assessment of a set of soft skills, with particular attention to the use of an automated prototype for the assessment of Critical

Thinking. We will conclude by outlining two research questions that have guided our research work. The use of open-ended questions represents a consistent approach for the observation of critical reasoning, because it allows students to freely articulate their thoughts, highlighting the analysis and synthesis strategies they implement (Ku, 2009; Liu et al., 2014). Unlike closed-ended tests, open-ended questions:

- promote freedom of expression: students are faced with a non-binding assignment, forced to organize and structure their thoughts to produce a coherent text;
- require analysis and evaluation: the lack of explicit prompts (such as response options) implies the need to independently use interpretation, argumentation and source evaluation skills;
- highlight metacognitive processes: those who answer an open-ended question can show how they reach a conclusion, revealing a part of that "process" that, in multiple-choice tests, often remains invisible.

This approach responds to the needs that emerged from the criticisms of standardized tests, where it is possible to obtain a positive score even without real reasoning (Halpern, 2007; Rear, 2019). However, the manual analysis of open-ended answers raises problems of inter-subjectivity (agreement between raters) and time costs (Ku, 2009), especially in contexts with a high number of students. To overcome these limitations, in recent years prototypes of automatic evaluation (Natural Language Processing, NLP) have been developed that, through text-mining algorithms, try to identify the typical patterns of critical reasoning (Poce et al., 2019b). The reference text describes one of these prototypes, which analyzes four macro-indicators (e.g., basic linguistic skills, relevance, importance and novelty) to quantify, in an automated way, the level of Critical Thinking in a written text (Poce, 2017). Although Critical Thinking is relevant in all disciplines, its explicit integration in educational paths is often associated with the humanities and social sciences, where argumentation and debate play a crucial role (Moore, 2013). On the contrary, in the so-called "hard sciences" (engineering, physics, chemistry, mathematics) there is a tendency to sometimes believe that the acquisition of technical knowledge is the core of education, leaving transversal skills, including CT, in the background (Kuhn, 2019). In reality, rapid technological and industrial evolution has made it increasingly clear that analytical, problem-solving and critical evaluation skills are also essential for engineers and scientists, who are called upon to manage:

- innovative processes: research and development of technological solutions require the ability to formulate hypotheses, evaluate results and constantly review one's approach;
- complex decisions in short time: in the engineering field, evaluation errors can have significant consequences in terms of safety and economic costs;
- interdisciplinary collaboration: large technological challenges often involve mixed teams (e.g. engineers, designers, managers), which require communication and critical comparison skills to manage information from different sources and visions.

For these reasons, engineering faculties and more generally STEM disciplines (Science, Technology, Engineering, Mathematics) have begun to understand the importance of integrating teaching and assessment activities aimed at promoting Critical Thinking (Ennis & Weir, 1985; Tiruneh et al., 2014).

3. The group of analysis

As part of a broader experimentation on soft skills, a group of engineering students participated in a workshop devoted to the development of transversal skills, with a particular focus on Critical Thinking. In order to carry out the assessment, it was decided to adopt a platform that combines manual and an automated assessment taking based on NLP and Deep Learning techniques. The key points of the workshop are summarized below.

In the introduction to Critical Thinking and Soft Skills, students were introduced to the importance of soft skills in the context of engineering 4.0.

Examples were provided of professional situations where CT is essential (project development, risk management, reliability assessment of technical sources).

Participants discussed definitions and components of CT, in line with the taxonomy presented by the *Delphi Report* (Facione, 1990). The four macro-indicators (basic linguistic skills, relevance, importance and novelty) were explored in depth, already (Poce, 2017; Poce et al., 2019a). Students were then asked some open-ended questions focused on engineering problems, inviting them to provide reasoned answers. The answers were collected both for manual analysis (conducted by expert CT teachers) and for automatic analysis, through a prototype that uses Wikipedia references and concept extraction modules to calculate the scores for each macro-indicator. This dual approach (manual and automated) allows to verify the agreement between human raters and the algorithm, facilitating the calibration of NLP tools. At the same time, it allows to provide students with immediate feedback, an aspect particularly appreciated in university courses with a high number of enrollments (Mao et al., 2018). This investigation arises from the need to deepen the effectiveness of the use of open questions and automated assessment tools to measure Critical Thinking in the engineering field. To this end, two main research questions were defined:

1. RQ1: what is the level of agreement between the manual assessment (expert humans) and the automatic assessment (NLP prototype) of Critical Thinking in the open-ended answers provided by engineering students? The aim is to understand whether the automatic analysis, based on the four macro-indicators (basic language skill, relevance, importance, novelty), is able to get close to the experts' judgments.
2. RQ2: to what extent does participation in a CT workshop foster the development of critical skills also in the STEM field, measured through open-ended responses? This question aims to explore whether a short training intervention (workshop) can represent a way of detection of critical thinking in the way in which engineering students formulate, evaluate and justify their arguments.

Before the actual start of the assessment test, students were pre-registered on the MYIntellect platform, creating simplified access credentials for all participants and direct access to the test after logging in. This operation was necessary both to make students with the digital environment and to allow the research group to set up effective monitoring of results without being influenced by technical issues that the use of the platform could cause. Following this, a face-to-face class was organized whose main purpose was twofold:

- delete any technological gap: some participants, in fact, had never used a Learning Management System (LMS) before. During the class, the basic steps to access the platform, insert their own papers and view feedback were shown;
- introducing cross-curricular skills, with a specific focus on Critical Thinking: the class illustrated the key concepts of CT (definitions, importance in academic and

professional fields), also providing practical examples of how these skills can be crucial in different disciplines and life contexts.

At the end of the pre-registration and training phase, the sample on which to conduct the analysis was defined. It consisted of 45 students, of which 16 were female (approximately 35.6% of the total) and 29 were male (approximately 64.4%) (Figure 1).

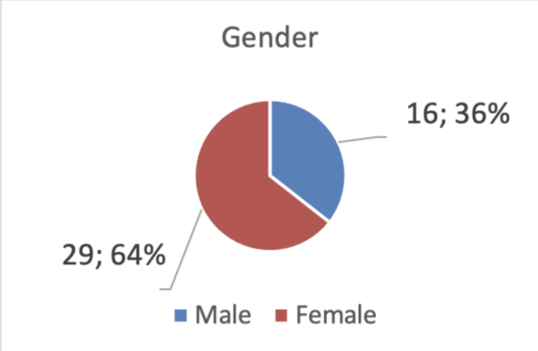

Figure 1. Gender distribution.

After the lecture, participants had access to the evaluation module within MYIntellect. They were offered an open test based on a short text by Descartes, taken from the *Discourse on the Method*. Part one: “But as soon as I had acquired some general notions of physics, and began to test them in some particular problem, I understood to what point they could lead and how much they differed from the principles that have been used up to now, I believed that I could not keep them hidden without sinning gravely against the norm that obliges us to promote as much as we can the general good of all men”. The test consisted of two open questions, to be uploaded to the Moodle platform:

3. paraphrase the passage: in this section, students were asked to rephrase the content of the text in a simpler and more linear way, thus revealing their ability to understand the literal meaning and reorganize it in their own words;
4. comment on the passage: what does the author mean? This part aimed to evaluate the ability of critical analysis: students had to interpret Descartes’ intentions, contextualize his reflection and, if possible, correlate the main ideas with contemporary or personal situations.

Students had a time limit (45 minutes) to read the text, carry out the paraphrase and develop the critical commentary. At the end, they were asked to upload the paper in text format (or PDF) directly to Moodle. At this stage, MYIntellect recorded the date and time of delivery of each participant. The research group used an assessment grid specifically designed for the acquisition of Critical Thinking knowledge and skills. This grid, validated in previous experiments (Poce, 2017; Poce et al., 2019a), presents six fundamental macro-indicators:

- use of language: evaluates spelling and morphosyntactic correctness and lexical propriety, with particular attention to the use of technical-scientific language;
- justification/argument: tests the ability to structure a thesis and to coherently argue theoretical, methodological and didactic choices;
- relevance: measures how closely the content developed adheres to the proposed outline, evaluating the understanding of the various points characterising the text or research project;
- importance: indicates how much the student demonstrates knowledge of the topic

being discussed, mentioning the main relevant aspects appropriately;

- critical evaluation: it assesses the ability to critically rework the available material (text, bibliographical sources), expressing coherent and in-depth personal reflections;
- novelty: estimate the presence of original information or ideas, as well as innovative solutions.

Each of the six macro-indicators is evaluated on a scale of 1 to 5 points, contributing to the integrated measurement of Critical Thinking. In particular, the descriptors provided by the grid help the evaluators to accurately detect the level of acquisition of each individual skill, from the use of an adequate vocabulary to the ability to elaborate innovative observations. In addition to the manual evaluation conducted by experts, an automated prototype (based on Natural Language Processing techniques) was integrated into the platform, with the aim of providing immediate feedback on the level of linguistic complexity, relevance and originality of the answers. Subsequently, the scores from manual and automatic evaluation were compared to validate the reliability of the system and to offer students cross-feedback. The automated assessment was conducted using a deep learning-based AI that had been trained on a corpus of written responses and evaluations originally performed on the same test in the humanities domain. The AI model draws on linguistic and semantic features derived from this corpus to approximate human judgment[2]. Using a platform like MYIntellect, built on Moodle, has greatly simplified the organization and collection of data. Students have benefited from:

- centralized access to resources: Descartes' text, guidelines for carrying out the exam, evaluation rubrics, technical assistance forum;
- timely feedback: manual correction by multiple human experts, combined with automated feedback, highlighted both correct understanding of the text and critical reasoning skills;
- delivery tracking: delivery dates and times were recorded, allowing for timing to be verified and learning progress to be assessed.

The sample of 45 students represents an adequate size for a first pilot activity, useful for collecting useful data for subsequent analyses. This group will allow us to explore possible gender differences in the use of Critical Thinking, as well as to detect possible correlations between personal background and performance on the test.

4. Results

In Figure 2 some examples of the human evaluation are presented.

From these data, it appears that the average scores assigned by human evaluators tend to be relatively high, often above 4.0 on a 1-5 scale. In general, the "Use of language" indicator shows that teachers have noted a good command of both vocabulary and syntax, with scores mostly ranging between 4 and 4.5 and occasionally reaching 5 in particularly outstanding cases. As for "Importance" and "Relevance", the analyzed texts display an

[2] All details regarding the training protocol, validation steps, and the specific methods adopted for corpus creation and annotation can be found in Poce et al. (2019a).

adequate knowledge of the subject matter and strong alignment with the instructions provided; many of these written responses earn scores approaching 5. “Argumentation” typically reflects a coherent, well-structured approach supported by relevant examples or evidence. However, “Critical evaluation” sometimes registers slightly lower values, such as 3.5 or 4, indicating that certain submissions remain at a descriptive level without offering in-depth analysis. Regarding “Novelty”, the ability to include additional information or original ideas varies, but overall the scores tend to fall into a mid-to-high range (4-5).

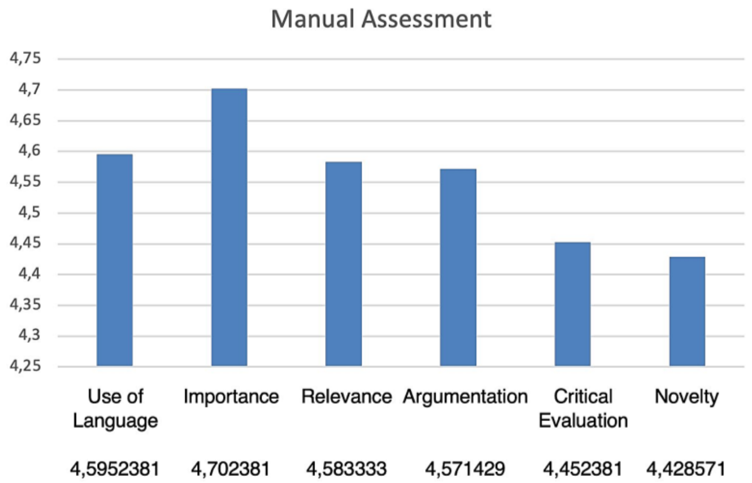

Figure 2. Manual assessment.

The first expert review has confirmed that the submissions demonstrate a generally good level of Critical Thinking. Nevertheless, some students have shown difficulties in linking theoretical concepts to real-life situations or in proposing truly innovative perspectives. In those instances, the overall average dropped to around 3.5-3.9, on the other hand, presents the scores assigned by the automated prototype. Below are a few examples.

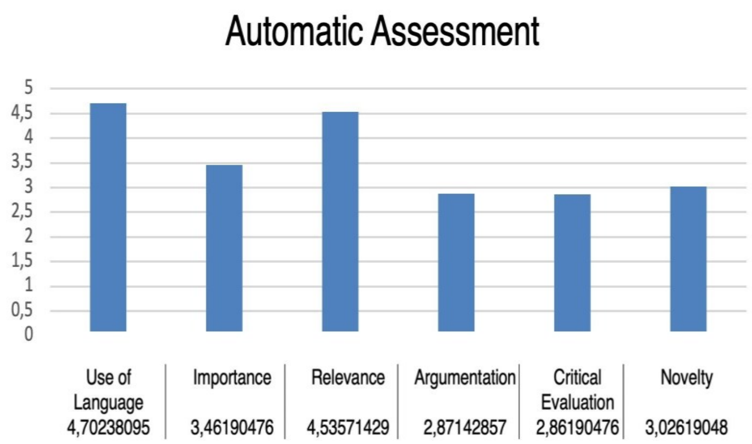

Figure 3. Automatic Assessment

The main results (Figure 3) highlight that:

- “Use of language”: the automatic system tends to provide scores of 4 to 5, in line with human judgments. This is due to the algorithms’ ability to recognize spelling errors, punctuation and some basic syntactic structures;

- "Importance" and "Relevance": scores are lower on average than human ratings. The algorithm probably relies on keywords and textual similarity, without fully capturing conceptual depth;
- "Argumentation", "Critical evaluation", "Novelty": these indicators show the main deviations. The system often assigns values from 1.5 to 3 where human evaluators tended to give scores above 4.

The final averages shown in the right column reveal how, in most cases, the automatic score is 0.5 to 1.0 points lower than the human evaluation. This trend suggests that the prototype consistently penalizes papers with "non-standardized" structures or arguments.

Following the observation of significant differences between the scores assigned by experts and those generated automatically, a second analysis was carried out focusing on the texts with the greatest discrepancies. In this phase, the so-called "problematic" papers were re-read to identify those elements that human evaluators had appreciated but the algorithm had been unable to recognize. Particular attention was then paid to linguistic features such as the use of synonyms, infrequent constructions, metaphors, or complex subordinate clauses, since these factors may have posed difficulties for parsing and scoring procedures. The analysis also examined how the system handled the presence of original ideas or conceptual links not explicitly indicated by keywords – for instance, interdisciplinary or implicit references – that are often hard to detect through a simple frequency or co-occurrence analysis. The in-depth review of these situations revealed that, in certain cases, discursive coherence built around personal reflections or examples not traceable to standard textual indicators was penalized when the algorithm failed to identify canonical argumentative sequences; similarly, several passages rich in reflection and critical insight were not recognized as such because they were formulated narratively instead of following more explicit models of expression. Another apparent limitation emerged in the realm of "Novelty", where the absence of codified introductory formulas often led the system to overlook additional information or genuinely innovative ideas. These findings gave rise to various hypotheses for improvement, including expanding the repertoire of synonyms, connectives, and rhetorical structures the prototype can manage, as well as integrating advanced semantic models (such as Word Embeddings or next-generation Language Models) to move beyond an evaluation based solely on terminological matches. A further step could involve detecting indirect signals of originality by analyzing hypothetical statements or sections of text that reveal genuinely autonomous thinking, even in the absence of conventional keywords.

5. Discussion

The findings presented in this study clearly highlight both the potential and the limitations of an automated system for assessing Critical Thinking. When comparing the scores assigned by human evaluators with those generated by the NLP prototype, a varied picture emerges: on one hand, the algorithm proves fairly accurate in identifying basic linguistic aspects (appropriate punctuation, spelling correction, syntactic complexity); on the other, it shows more pronounced difficulties with indicators that have a strong interpretive component, such as "Argumentation," "Critical Evaluation", and "Novelty". In these instances, its ability to capture semantic nuances, logical structures, and original ideas still appears limited. This gap often translates into an average difference of about 0.5-1 point to the disadvantage of the automated assessment compared to the human evaluation. Within the framework of the two research questions (RQs) guiding the study, several conclusions can be drawn. In response to RQ1, namely "To what extent is there agreement between manual (expert human) evaluation and automated (NLP prototype) assessment of Critical Thinking?", the data suggest that the correspondence is good only for more formal aspects (such as linguistic accuracy) and declines markedly when dimensions related to critical elaboration and originality of argumentation come into play. This finding confirms that while the prototype is effective as a support tool in the early stages of grading, it is not yet capable of matching human judgment in terms of depth and interpretive richness. Regarding RQ2 – "To what extent does participation in a CT workshop foster the development of critical skills also in the STEM field, measured through open-ended responses?" – the question aims to explore whether a short training intervention can serve as an avenue for detecting critical thinking in the way engineering students formulate, evaluate, and justify their arguments. The results show a tangible improvement in the texts produced after the training program, with generally higher scores in both human assessments and, to a lesser extent, automated evaluations. This indicates that a focused instructional approach (e.g., specific exercises and targeted feedback) can positively influence the development of critical thinking skills even in technical disciplines, although the NLP system still struggles to capture the more creative or complex elements introduced by students in their work. Despite these discrepancies, the utility of an automated scoring system remains indisputable in contexts with a large number of students or in teaching activities that require continuous progress monitoring. The significant reduction in tutors workload, the possibility of obtaining immediate feedback, and the consistency in assigning certain scores (particularly with respect to language usage) are strengths that justify the adoption of such technologies. Moreover, having an initial "objective" judgment can encourage students to carefully revisit their written work, devoting special attention to formal aspects and to a clearer and more linear structure. It is important to remember, however, that Critical Thinking often manifests itself through complex elaborations characterized by interdisciplinary references, personal examples, and sometimes non-linear argument structures. Precisely for this reason, the system appears lacking in its ability to recognize highly reflective passages conveyed through metaphors, analogies, or implicit references. In our study, this shortcoming led to a risk of underestimating high-quality papers, assigning them modest scores simply because they did not adhere to standard textual patterns. Nevertheless, the findings confirm that an automated approach can serve as a valuable support tool when integrated with careful expert supervision, especially for large-scale assessments or blended learning contexts. By continuing to refine the algorithm's semantic models and expanding the range of textual features it can identify, future iterations of automated scoring systems could come closer to evaluating complex reasoning processes that extend beyond conventional discourse structures. This combination of technology and human judgment points toward a promising horizon in which the benefits of speed and consistency in evaluation can be aligned with a more nuanced recognition of critical and creative thinking.

Reference list

Barton, D., Farrell, D., & Mourshed, M. (2013, January 1). *Education to employment: Designing a system that works*. McKinsey Center for Government.

Becker, G. S. (1975). *Human capital*. University of Chicago.

Bettiol, M., & Micelli, S. (2020). Competences for industry 4.0: The case of dual system in the US. *Industry and Higher Education*, *34*(5), 321–332.

Blömeke, S., Gustafsson, J. E., & Shavina, E. (2017). What is the place of creativity in a measurement of twenty-first-century skills? *Educational Research and Evaluation*, *23*(5-6), 230–234.

Bughin, J., Hazan E., Lund, S., Dahlström. P., Wiesinger, A., & Subramaniam, A. (2018, May 23). *Skill shift: Automation and the future of the workforce*. https://www.mckinsey.com/featured-insights/future-of-work/skill-shift-automation-and-the-future-of-the-workforce

Capelli, P. (2015). Skill gaps, skill shortages, and skill mismatches: Evidence and arguments for the US. *ILR Review*, *68*(2), 251–290.

Cobo, C. (2013). Skills for innovation: Envisioning an education that prepares for the changing world. *Curriculum Journal*, *24*(1), 67–85.

Deming, D. J. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics*, *132*(4), 1593–1640.

Earley, P. C., & Ang, S. (2003). *Cultural intelligence: Individual interactions across cultures*. Stanford University.

Ennis, R. H., & Weir, E. (1985). *The Ennis-Weir critical thinking essay test*. Midwest Publications.

European Commision. EC. (2020). *European skills agenda for sustainable competitiveness, social fairness and resilience*. https://ec.europa.eu/commission/presscorner/detail/en/ip_20_1196

Eurostat (2025) EU labour market - quarterly statistics https://ec.europa.eu/eurostat/statistics-explained/index.php?title=EU_labour_market_-_quarterly_statistics

Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction. The Delphi Report*. California Academic.

Gilbert, J. (2012). Soft skills and the future of work. *On the Horizon*, *20*(4), 322–326.

Goleman, D. (1995). *Emotional intelligence: Why it can matter more than IQ*. Bantam Books.

Grugulis, I., & Vincent, S. (2009). Whose skill is it anyway? Soft skills and polarization. *Work, Employment and Society*, *23*(4), 597–615.

Halpern, D. F. (2007). *Halpern critical thinking assessment using everyday situations: Background and scoring standards*. Claremont McKenna College.

Heckman, J. J., & Kautz, T. (2012). Hard evidence on soft skills. *Labor Economics*, *19*(4), 451–464.

Italian National Institute of Statistics. Istat. (2022). *Rapporto annuale 2022. La situazione del paese*. Istat. https://www.istat.it/storage/rapporto-annuale/2022/Rapporto_Annuale_2022.pdf

Klaus, P. (2010). Communication breakdown. *California Job Journal*, *28*(1), 1–9.

Ku, K. Y. L. (2009). Assessing students' critical thinking performance: Urging for measurements using multi-response format. *Thinking Skills and Creativity*, *4*(1), 70–76.

Kuhn, D. (2019). Critical Thinking as discourse. *Human Development*, *62*(3), 146–164.

Lam, Q. K. H. (2019). Lifelong learning in Asia: The changing nature of work and skills. *Asia Pacific Education Review*, *20*(4), 563–577.

LinkedIn Learning. (2023). *Workplace learning report*. https://learning.linkedin.com/resources/workplace-learning-report-2023

Liu, O. L., Frankel, L., & Roohr, K. C. (2014). Assessing critical thinking in higher education: Current state and directions for next-generation assessment. *ETS Research Report Series*, *2014*(1), 1–23.

Luthans, F., Youssef, C. M., & Avolio, B. J. (2007). *Psychological capital: Developing the human competitive edge*. Oxford University.

Mao, L., Liu, O. L., Roohr, K., Belur, V., Mulholland, M., Lee, H. S., & Pallant, A. (2018). Validation of automated scoring for a formative assessment that employs scientific argumentation. *Educational Assessment*, *23*(2), 121–138.

Moore, T. (2013). Critical thinking: Seven definitions in search of a concept. *Studies in Higher Education*, *38*(4), 506–522.

Organization of Economic Co-operationa and Development. OECD (2019). *Skills outlook 2019: Thriving in a digital world*. OECD Publishing.

Poce, A. (2017). *Verba sequentur. Thinking and writing for a critical development of skills in secondary school*. FrancoAngeli.

Poce, A., Amenduni, F., De Medio, C., & Re, M. R. (2019a). Road to critical thinking automatic assessment: A pilot study. *Form@re. Open Journal per la Formazione in Rete*, *19*(3), 60–72.

Poce, A., De Medio, C., Amenduni, F., Re, M. R. (2019b). Critical thinking assessment: A first approach to the automatic evaluation. In *2019 18th International Conference on Information Technology Based Higher Education and Training (ITHET)* (pp. 1-8). IEEE.

Rear, D. (2019). One size fits all? The limitations of standardized assessment in critical thinking. *Assessment & Evaluation in Higher Education*, *44*(5), 664–675.

Robles, M. M. (2012). Executive perceptions of the top 10 soft skills needed in today's workplace. *Business Communication Quarterly*, *75*(4), 453–465.

Roselina, S. (2009). Soft skills at the Malaysian institutes of higher learning. *Asia Pacific Education Review*, *10*(3), 309–315.

Schulz, B. (2008). The importance of soft skills: Education beyond academic knowledge. *NAWA Journal of Language and Communication*, *2*(1), 146–154.

Tiruneh, D. T., Verburgh, A., & Elen, J. (2014). Effectiveness of critical thinking instruction in higher education: A systematic review of intervention studies. *Higher Education Studies*, *4*(1), 1–17.

Trilling, B., & Fadel, C. (2009). *21st century skills: Learning for life in our times*. Jossey-Bass.

Unioncamere & Anpal. (2024). *Previsioni dei fabbisogni occupazionali e professionali in Italia a medio termine (2024-2028). Scenari per l'orientamento e la programmazione della formazione*.

https://excelsior.unioncamere.net/sites/default/files/pubblicazioni/2024/report_previsivo_2024-28.pdf

Van Laar, E., Van Deursen, A. J., Van Dijk, J. A., & De Haan, J. (2017). The relationship between 21st century skills and digital skills: A systematic literature review. *Computers in Human Behavior*, *72*, 577–588.

Weber, M. R., Finley, D. A., Crawford, A., & Rivera, D. (2009). An exploratory study identifying soft skill competencies in entry-level managers. *Tourism and Hospitality Research*, *9*(4), 353–361.

World Economic Forum. WEF. (2020). *The future of jobs report*. https://www.weforum.org/publications/the-future-of-jobs-report-2020/

Zehr, M. A. (1998). Hard lessons about soft skills. *Education Week*, *18*(10), 28–29.