Multidisciplinary Journal for Education
Social and Technological Sciences

https://doi.org/10.4995/muse.2022.17908
e-ISSN: 2341-2593

Working on Critical Thinking skills using the computer lab works of an Engineering subject

Juan Giner-Navarro[1], Águeda Sonseca[1], José Martínez-Casas[1], Javier Carballeira[1]

[1]Department of Mechanical Engineering and Materials Science, Universitat Politècnica de València, 5E Building, Camí de Vera, s/n, 46022, Valencia, Spain. Emails: juanginer@upv.es; agsonol@upvnet.upv.es; jomarc12@mcm.upv.es; jacarmo@mcm.upv.es (corresponding author)

Received: 13 July 2022; Accepted: 22 August 2022; Published: October 2022

Abstract

A practical intervention was devised in order to promote and to assess Critical Thinking in undergraduate students of Engineering. First, in collaboration with specialists in education, the concept of Critical Thinking, and its skills and dispositions on which to focus on, was determined. Second, a teaching strategy was designed in order to be as effective as possible, regarding the acquisition of the competence by students, considering the limitations of the intervention. This strategy took advantage of the computer lab sessions where discussion and questioning encouraged the development of Critical Thinking. Then, the instruments to assess the acquired skills and attitudes of the students were developed. Finally, a quantitative analysis of the results was conducted in order to evaluate the validity and reliability of the strategy. This paper presents a full description of the intervention carried on for two years. Besides the desired effects on the students' performance, some conclusions regarding the development of appropriate instruments to deal with a large group of students are drawn. This intervention has proven to be effective in order to help the students to develop their Critical Thinking skills, and it is particularly suitable for large groups.

Keywords: Critical Thinking; learning assessment; teaching strategies; engineering education; engineering students.

To cite this article: Giner-Navarro, J., Sonseca, A., Martínez-Casas, J., Carballeira, J. (2022). Working on Critical Thinking skills using the computer lab works of an Engineering subject. Multidisciplinary Journal for Education, Social and Technological Sciences, 9(2), 23-45. https://doi.org/10.4995/muse.2022.17908

https://doi.org/10.4995/muse.2022.17908
e-ISSN: 2341-2593

1. Introduction

Engineering problems in real-life are usually ill-structured and complex because they possess "conflicting goals, multiple solution methods, non-engineering success standards, non-engineering constraints, unanticipated problems, distributed knowledge, collaborative activity systems, the importance of experience, and collaborative activity that rely on multiple forms of problem representation" (Jonassen et al. 2006). It has also been defined as "a purposeful, reflective judgment, which manifests itself in reasoned consideration of evidence, context, methods, standards, and conceptualization in deciding what to believe or what to do" (Facione 1990). Critical Thinking (CT) skills such as analysis, evaluation, interpretation and self-regulation are essential to develop the engineering judgement through solid decision-making and problem-solving abilities in order to address these problems (Pierce et al. 2013; Baytiyeh and Naja 2017). Therefore, CT has always been considered as a key competence for engineers and, in general, a crucial learning outcome for higher education (Stassen et al. 2011; Association of American Colleges and Universities (AACU) 2009; Hart Research Associates 2013). This competence is hence highly valued by employers to optimize recruitment (Andrews and Higson 2008).

Furthermore, these skills are recognised to be critical in preparing the students for increasing social responsibility and their own individual career development (Kelly 2001; Rieckmann 2012). Sustainable development problems such as climate change, poverty and resource scarcity share those characteristics with engineering problems, which is what makes them really challenging for engineering students (Lönngren and Svanström 2015; Llopis et al. 2022). So, having well-developed CT skills is not only important for students to get a good job, but also to become better professionals and citizens (Paul 1995).

Developing CT skills and dispositions can improve the students' learning, moreover when there is such an amount of information available through the internet (Naimpally 2012; Douglas *et al.* 2018). The relation between CT and problem-solving abilities is especially significant in engineering (Itabashi-Campbell 2011). Therefore, most of the frameworks in engineering education include these skills, together with some other generic competencies (Young and Chapman 2010; Sursock and Smidt 2010).

https://doi.org/10.4995/muse.2022.17908
e-ISSN: 2341-2593

Nevertheless, several studies emphasize that students need to be able to evaluate and solve problems rather than memorize and deploy algorithms to meet workplace demands (Marshall and Horton 2011; Dunne 2015), and there is evidence that this need is not presently being met. Prior research endorsed that CT skills appear to be lacking among fresh engineering graduates (Jackson 2010; Schulz 2008). Some interviews made by Dunne (2015) reported that, in general, employers contend that university graduates CT skills required in the workplace. In this line, the recent CRITHINKEDU European project on the CT skills and dispositions needed in different professional fields, states that there is a gap between the technical skills and the CT skills of the graduates (Dominguez *et al.* 2018). It was emphasised by the employers that, while students have strong technical skills, their CT skills have not been developed sufficiently. This gap was also recognised some years ago in the governing bodies of the authors' university – Technical University of Valencia, Spain (UPV). So, an institutional project to foster innovative learning and teaching activities related to generic competencies, including CT, was implemented (UPV 2020). These findings demonstrate that CT is still a pivoting and demanded skill for students to get a good job, leading to the need to incorporate CT into traditional teaching methodologies in order to enhance their abilities to identify and solve problems within the professional context.

One main issue to deal with generic competencies is their assessment (Zlatkin-Troitschanskaia *et al.* 2015). Engineering academics feel confident about the methods to teach and to evaluate technical skills, but they have more trouble to do the same with CT, in part because there is a wide range of understanding of the CT skills. Engineering academics have more difficulties to define what CT is than academics in non-technical disciplines (Ahern 2012). This leads to the use of a variety of pedagogical strategies for its promotion and assessment (Michaluk 2016; Guerra and Holgaard 2016; Hagerty and Rockaway 2012), and this variety makes it difficult to measure or monitor their effectiveness (Ahern 2019).

Furthermore, we decided carefully which skills and dispositions of CT to focus on, as there were some limitations to our intervention. Fortunately, our university took this first step for us. Within the institutional project mentioned above, and in collaboration with specialists in education, a set of definitions and learning outcomes associated with the generic competencies was given.

https://doi.org/10.4995/muse.2022.17908
e-ISSN: 2341-2593

This definition of CT includes the ability to apply logical and rational processes to analyze the components of an issue, and to think creatively to generate innovative solutions. It also states that developing a disposition toward CT goes along with being interested in the foundations on which the ideas, actions and judgments, both their own and those of others, are based. CT goes beyond the skills of logical analysis, since it involves questioning the underlying assumptions in our habitual ways of thinking and acting. It refers to being prepared to think and do differently, based on that critical reasoning. Consequently, we will say that the students have developed it to the extent that they question themselves about things and are interested in the foundations of the ideas, actions, assessments, and judgments. According to this definition, the main CT skills we focused on are interpretation, analysis and evaluation, and the main disposition to promote is willingness to questioning.

From the authors' point of view, developing these skills can be very helpful for the students to deal with technical subjects in which they have to put their cognitive capacities into play. So, a positive correlation between CT level and students' performance should be expected. Therefore, the main goals of our intervention were: to promote and assess CT; to develop appropriate evaluation instruments to deal with a large group of students; and to check if there was a positive correlation between CT and students' performance.

To reach these goals, we had to decide how explicit we would be about teaching CT. We could have used one of the three following options: a General approach and teach CT separately from the subject matter; an Infusion approach and teach CT embedded in the subject matter, or an Immersion approach and teach CT "involving students in a dialogue where they are prompted to consider, analyze and evaluate different points of view" (Angeli and Valanides 2009). Then, in order to promote the CT skills and dispositions, the learning activities that better fit our approach were planned: problem-based learning, case-based studies, lecture discussion, open-ended questions, role-play, essays, interviews, observations, and so on.

A careful assessment of the acquired skills and attitudes must be designed to measure the effectiveness of the teaching strategy (Liu *et al.* 2014). Conventional tests within the subjects not only provide ease and timeliness in quantifying students' knowledge, but they also come at the expense of helping them become better critical thinkers (Shuman *et al.* 2004). Nevertheless, not to

https://doi.org/10.4995/muse.2022.17908
e-ISSN: 2341-2593

have separate instruments for assessing the students' CT prevent educators from gaining insight to students' learning processes (Birenbaum 1996). These instruments will be more effective when designed for learning (Angelo and Cross 1993), trying to challenge the students to learn by constructing new meanings, developing strategies, thinking critically about solution implications, and evaluating their own thinking during the task (McCormick *et al.* 2014). The idea is to use the assessment to interpret students' reasoning and ways of thinking critically about the problem. Depending on the approach and the activities, the instruments for assessing the students may be different. The use of more than one measure is strongly recommended to give consistency to the further data analysis (Ahern 2019).

Our approach for the assessment of CT is based on the knowledge that students must really understand the concepts to pass the exams in courses for such challenging degrees as Engineering ones. From this approach, we would state that the instrument is practical for our purposes when there is a good agreement between the CT grading and the students' performance in the subject, i.e. students with better grades in the subject should also have a higher score in CT. Nevertheless, we will assume that most of our students will earn grade high on CT, as they are already students who have undergone a very selective process to access to this degree. However, CT are not the only skills and dispositions necessary to be successful. Therefore, a direct relationship between CT grading and the students' performance in the subject should not necessarily be expected.

2. Method

The intervention was implemented in the second year of the bachelor's degree in Aerospace Engineering. It was carried on during 2018 and 2019 in Mechanics, a subject related to kinematics and dynamics of rigid bodies and mechanical systems. The intervention involved 116 students in 2018 and 103 in 2019, divided into six subgroups for the lab sessions. The global demographics: 21% were female, 74% were freshman, 22% were sophomores and 4% were juniors, 45% were in High Academic Performance groups.

https://doi.org/10.4995/muse.2022.17908
e-ISSN: 2341-2593

2.1. Teaching strategy

Mechanics is traditionally considered one of the toughest subjects of the degree by most of the students (evaluations compared with other subjects of a technical nature certify this) since it requires developing knowledge related to complex mathematical and physical concepts. The objective of the teaching strategy proposed is to try to find a way in which it wouldn't be even more difficult while we work CT. Both the technical nature of the subject compels to a rethinking of the way of teaching CT (Ahern 2012), as does the large number of 100 students. Once we clarified what skills and dispositions need to be taught and considered the practical limitations described, we decided to use the computer lab works to develop CT using an Immersion approach. This approach allows teaching CT within the normal progression of the subject matter, without causing readjustment to the contents of the teaching guide (which is already aligned with the number of hours available to the subject).

In our course, the theory sessions are carried out dividing the total group into two: the High Academic Performance (HAP) group, with about 45% of the students, and the remaining 55% (no High Academic Performance, nHAP), so that both have more than 50 students each. The design of any intervention associated with the development of CT for such large groups was not feasible if we pretend to complete the teaching guide. In addition, each of these theory groups are separated into three subgroups for the laboratory component of the course because the number of computers per laboratory is limited to 12. The students work in pairs along seven sessions (2 hours each) with the objective to solve different applied cases from which they had to extract kinematic and dynamic results. They make use of the commercial simulation software ADAMS/View© to do some exercises and practical demonstrations that help them to visualize and understand some of the main concepts in the subject, i.e. the software is used as a virtual lab.

These computer lab works are designed to be suitable to work on CT by putting the CT skills and dispositions into play. Using software with several alternatives to solve the exercises forces the students to analyze, understand and evaluate the information they are given, and to think critically about the choices they make to accomplish their goals. Moreover, students can check most of their results using analytical methods they have learnt in the theory sessions with real data provided. So, a comparison between what they should expect and what they get from the software is available, which allows them to be critical users of numerical tools in contrast with real experiments.

https://doi.org/10.4995/muse.2022.17908
e-ISSN: 2341-2593

2.2. Intervention design

The fundamental advantage of having limited capacity to a maximum of 24 students per lab session is that the practical application of the intervention becomes feasible. Each pair is given a different exercise with geometry and inertial data differing from the reference problem developed and solved in the available tutorial. The intervention is organised and planned in three phases that last in total the 2 hours of each session. The distribution is as follows:

Phase 1: Introduction to the lab work. During the first 20 minutes, the instructor introduced the main goals of the session, detailed the problem that the students faced and how to approach it through the simulation software, in line with the physical conceptual framework previously worked on during the theory sessions. The students had the support of a tutorial to help them in the use of the software. Key indications for the most general difficulties in the software implementation (based on the experience of previous years) were also given.

Phase 2: Segmented debates. The students had 80 min to implement the dynamic model, and they could ask for the instructor's support twice as maximum per work pair, forcing them to be self-sufficient, to try to overcome the problems they encountered with the available means, and to critically select the questions they would later ask the lecturer. The lab works are designed so that the students face one or two critical moments that require decisions on their criteria sustained on their physical background. The instructor takes advantage of these critical-based steps to work the competence within reduced groups of two or three work pairs separately in order to have a more direct treatment with the students. The lecturer tries to compel each student to participate giving their critical opinion focusing the critical step to make progress in the dynamic modelling (Rodriguez-Largacha 2014). In this way, a proper interpretation of results, that may raise to substantial doubts, is ensured. Some of the reasoning that emerges might be really interesting thus, the instructor takes note of each student interaction through observation. In the following subsection, we describe the assessment instruments adopted and justify why this observation tool was only used in the first year for the assessment of CT.

Phase 3: Group intervention. The last 20 min of the session were reserved for a group analysis of the modelling process, the key decisions that each work pair made based on critical criteria, the results obtained, their comparison and the critical discernment of their plausibility based on the

https://doi.org/10.4995/muse.2022.17908
e-ISSN: 2341-2593

problems addressed in the theory sessions and reality experience. Again, the lecturer tried to get all the students, or at least all the pairs, to intervene in a debate. If some of the decisions made by each pair led to erroneous results, their reasons were analysed as was how to re-approach the dynamic model. That is why, in order not to condition lab scores with the development of CT, the students had a week to finish the lab work with the appropriate corrections, spurring the critical questioning of decisions previously made even after the lab session.

The lab work grade is the average of the seven sessions, evaluating the technical skills on the computer lab works. Since the requested results are numerical, an automated correction is available (assuming a 5% of tolerance margin) in the Intranet of the university website, through which the students complete each lab work.

2.3. Assessment instruments

To measure the effects of our intervention and the influence of CT on the students' performance, two assessment methods were devised in the first year of the intervention: a set of open-ended questions to be answered by the students in the last lab session, and the observation during the lab sessions. These methods would allow us to check if there was an improvement in the CT skills and dispositions of the students, and if there was a correlation between these skills and the performance in such a challenging subject as *Mechanics*. There were three instructors involved every year, so all these decisions were coordinated and discussed together in order to minimize the instructor bias.

Within the institutional project (UPV 2020) a rubric was developed to define the learning outcomes associated with this competence in the first years of bachelor's degree (see **Table 1**). Each of these learning outcomes are divided into four levels of achievement, with the corresponding numerical values interval: Not reached (D: 0–3.9 points), In Development (C: 4–5.9), Good/Adequate (B: 6–7.4, Excellent/Exemplary (A: 7.5–10). These intervals have worked well in other subjects when discriminating the degree of acquisition of other evaluated competencies, so it has been decided to keep them in this case.

https://doi.org/10.4995/muse.2022.17908
e-ISSN: 2341-2593

Adapting this rubric to our subject has required a critical rethinking of the assessment instruments. Two of these developed instruments were implemented in the first year experienced: 4-item open questionnaire and observation.

Multidisciplinary Journal for Education Social and Technological Sciences

https://doi.org/10.4995/muse.2022.17908
e-ISSN: 2341-2593

Table 1 Rubric to assess the learning outcomes of Critical Thinking for the two first years of the Bachelor´s degrees (UPV 2020)

Learning Outcome	D. Not reached	C. Developing	B. Good / adequate	A. Excellent / exemplary	Evidence
The student shows a critical attitude towards reality: the student wonders why things are happening	The student does not manifest any kind of critical spirit: he never questions the situation or the reality in which he lives. It assumes as true any information it receives	The student questions certain situations of the reality in which he lives. But it is incapable of issuing judgments and valuations of its own. Need the help of others to get answers	The student asks the question of things and investigates to get answers autonomously. But it is influenced by issuing its own judgments and assessments	The student reflects and investigates the why of things, and is able to find answers and argue them objectively	In exercises or problems with real proposals, include questions that invite the student to question: Is the result obtained in the previous section coherent (units, physical sense)? Could it have been solved in another way?
Detects inconsistencies or contradictions in other people's speech or in a text	The student is not able to detect inconsistencies or contradictions in a speech or text	The student is able to detect some inconsistencies but does not know how to explain why	The student detects inconsistencies and contradictions, and provides arguments to demonstrate the same	The student detects inconsistencies and contradictions, provides arguments, and coherently reformulates contradictory statements	To evaluate solving problem process; to suggest problems with multiple solutions; to justify the methodology and data used
Differences made from opinions, interpretations or valuations	The student shows a reflective attitude towards another people's discourse. Does not distinguish facts from opinions	The student normally distinguishes facts from opinions, but can accept judgments or decisions based on opinions	The student differences facts from opinions, interpretations or assessments in the arguments of others	The student questions judgments or decisions based on opinions, evaluations, etc. and detects fallacies and ambiguities	From press news or texts related to the subject, ask students to differentiate between objective facts, and interpretations of the author
Delves into a topic with logic and impartiality, contrasting information in reliable sources	The student is not able to delve into a subject. It uses a single source and does not contrast the information	The student resorts to diverse sources, but it does not verify the reliability of the same ones	The student consults different sources and contrasts the information of them to verify their reliability	The student consults reliable sources, contrasts the information and provides his personal assessment	Ask the students to do brief documentation work on a topic related to the subject, and to contribute the bibliography

https://doi.org/10.4995/muse.2022.17908
e-ISSN: 2341-2593

2.3. Open questionnaire

First, a competence whose nature demands a solid reasoning from the students when addressing an open problem requires an open-format assessment tool (Liu *et al.* 2014). It should allow the student to develop and synthesize a chain of ideas with a plot line that goes outside the framework of the acquired theoretical knowledge. However, we were dealing with large groups of students, so we needed an instrument that could be quickly completed by students and quickly assessed. Therefore, we opted for a short questionnaire with open questions as the appropriate trade-off between an open-format and time (Hong and Choi 2015).

The rubric shown in **Table 1** is a general-purpose instrument (UPV 2020), and so, the generic descriptors provided could not be directly applied in *Mechanics*. Thus, these descriptors had to be adapted and converted into open questions. These questions forced the students to analyze, understand and evaluate as a result of handling the software used in the lab sessions. This assessment instrument intends to assess the originality of their answers beyond the technical knowledge acquired, also proposing new cases as well as the advantages and disadvantages to implement them in the software.

The students had 20 minutes to respond to the four open questions. The questionnaire was delivered in the last session, following the format presented in **Table 2**. A set of guidelines was discussed to score the answers to each question in order to obtain a quantitative measurement of the students' CT level. The grade in the competence questionnaire was not included in the averaging of the course.

Table 2. Assessment instrument: open 4-item questionnaire

1. *"As the world's most famous and widely used Multibody Dynamics (MBD) software, ADAMS improves engineering efficiency and reduces product development costs by enabling early system-level design validation. Engineers can evaluate and manage the complex interactions between disciplines including motion, structures, actuation, and controls to better optimize product designs for performance, safety, and comfort. Along with extensive analysis capabilities, ADAMS is optimised for large-scale problems, taking advantage of high-performance computing environments."* With which affirmations of the text belonging to the ADAMS' brochure do you agree more and with which disagree? Argue your reasons.

https://doi.org/10.4995/muse.2022.17908
e-ISSN: 2341-2593

2. Explain which, in your opinion, are the advantages and disadvantages of using ADAMS/View© to solve a parabolic shooting problem with air friction.

3. What are, from your point of view, the factors that can cause the numerical results to move away from the real measures?

4. Which parameters do you consider most important to carry out a correct dynamic analysis? Why?

Item 1. It requires the critical evaluation of a promotional text of the software brochure, indicating with which statements students agree/disagree based on their experience. This item involves questioning the underlying assumptions and asks students to differentiate between objective facts and interpretations.

- **Rubric 1**. It is expected that the students question some statements from the brochure, such as the ease of use and flexibility of the software in some specific circumstances and especially "*ADAMS is optimised for large-scale problems*", since they have not seen such problems in any of the cases studied in the sessions.

Item 2. It raises the problem of parabolic shooting considering air friction, a question solved analytically in the theory sessions. This allows students to evaluate the problem-solving process for a situation with multiple solutions, requiring the justification of the methodology used.

- **Rubric 2**. We want to make students think whether, for a preliminary study, the time invested in a model in the software is profitable, being able to solve the problem analytically with a shorter procedure; the considerations about the complexity of modeling friction with air and other physical phenomena are also valued. This question opened a debate about the most optimal way to address a problem in which the aerodynamic forces completely condition the trajectory of the body under study.

Item 3. It focuses on the causes of divergences between the numerical results and the actual measurements, encouraging the students to take a critical perspective in the assumptions considered during the modelling process.

https://doi.org/10.4995/muse.2022.17908
e-ISSN: 2341-2593

- **Rubric 3**. It sought to expand on expectations versus reality regarding the energy dissipation mechanisms and predictable and unpredictable factors that deviate measurements with respect to the software results. Numerical precision of the software in the time integration can also be mentioned.

Item 4. The students are asked to choose the key parameters for solving dynamic cases, especially the ones that they consider crucial for the analysis.

- **Rubric 4**. The idea is that the students take control of the parameters that influence the dynamic study. The rigor and completeness of the answers have been well scored.

2.4. Observation

As the size of the groups was reduced substantially to a maximum of 24 students, an observation instrument was conducted only in 2018 as a direct measurement of the interactions and critical participation of the students in the lab sessions. Through debates applied to subgroups of two or three work pairs, the lecturer tried to encourage the students to question critically about whether, the tutorial approach was appropriate for their specific problem or they would have adopted other methodologies. Some students proposed alternatives that involved saving user time and/or allowing better visualization of the results and, hence, contrasting them in a more direct way.

3. Results

Data analyses are presented in this section to check the validity and the reliability of this implementation, to assess the influence of the CT intervention on the students' performance. **Table 3** shows the average score of all the students (*Total*), for the HAP and nHAP groups for each performance indicator, i.e. lab work, theory and global grades, together with the rates of the assessment based on observation (*Obs.*) only applied in 2018, and the open questionnaire instrument (*Quest.*) for 2018 and 2019. Average observation for CT showed that there were not statistically significant differences between the HAP and nHAP students (7.54 and 7.53, respectively). Regarding the open questionnaire, the CT rates are apportioned into the four items, among which the average

Multidisciplinary Journal for Education
Social and Technological Sciences

https://doi.org/10.4995/muse.2022.17908
e-ISSN: 2341-2593

of the item 4 in 2018 was significantly higher than the three previous questions and led us to check the origin of this deviation. It was decided to delete it after concluding that this item generated memory-based responses instead of reasoning-based ones as will be discussed later.

In 2018, there was a noticeable difference between HAP and nHAP students in the performance indicators around 1.3 points over 10, which was reflected with a deviation of 0.84 points for the CT assessment favorable to HAP students. There was detected a considerable increase of the average of the performance indicators in 2019, especially in the theory and global marks with more than 1 point. This increment led to a reduction in the difference between the HAP and nHAP average (0.62 points in the global marks), also shown in the CT rate with a shortened but still remarkable difference of 0.6 points favorable to HAP students. Hereby, the table demonstrates a preliminary trend in line with our expectations: students with better CT skills earn higher scores in the subject.

Table 3 Average of the questionnaire (indexed per items), lab work, theory and global marks for 2018 and 2019

		Item 1	Item 2	Item 3	Item 4	Questionnaire	Observation	Lab	Theory	Global
	Total	**6.53**	**6.20**	**6.55**	**7.30**	**6.64**	**7.54**	**7.39**	**4.56**	**5.32**
1st year	*HAP*	7.25	6.67	6.94	7.56	7.11	7.54	8.15	5.28	6.04
	n*HAP*	5.94	5.81	6.23	7.09	6.27	7.53	6.77	3.97	4.73
	Total	**6.77**	**6.44**	**6.57**	**-**	**6.32**	**-**	**7.73**	**5.77**	**6.35**
2nd year	*HAP*	7.03	6.57	6.79		6.79		8.11	6.13	6.67
	n*HAP*	6.58	6.33	5.75		6.19		7.35	5.44	6.05

To take a further step on the information that the results of the instruments used can provide, the correlation between performance and CT measured through both tested instruments can be visualised in the figures above. First, **Figure 1** allows to check that observation within the immersion approach experienced in the subject does not give any noticeable correlation or trend. The experience corresponding to 2018 made us quickly realize that the notes taken by the instructors and, finally, the score based on the observation unconsciously ended up depending more on how extrovert or not the students were than on their real abilities and dispositions regarding the competence under study. Subjectivity played a role that could hardly be avoided despite any established guideline or rubric,

https://doi.org/10.4995/muse.2022.17908
e-ISSN: 2341-2593

which increased with the number of instructors in lab works. Furthermore, under what criteria did we evaluate students who barely participated? Could we have created a situation that would have allowed us to have information to evaluate them? We had no time enough for individual interviews without interrupting the progress of the work in pairs during the lab session. So, the lack of correlation obtained could be expected and led us not to repeat this instrument in the second year.

Figure 2 compares the open questionnaire results implemented in 2018 (3 items, **Figure 2(a)**) and 2019 (4 items, **Figure 2(b)**) separately. The results for the 4-item questionnaire illustrate that, in general terms, scoring high in CT is indicative of a good performance in the subject; the trend line, which crossed at the origin, makes this more obvious. Similar correlations for the lab sessions and the theory exam were also obtained (not shown in this paper for conciseness). The suppression of the item 4 in the questionnaire allowed obtaining better results in the second year. Comparing **Figures 2(a)** and **(b)**, it is observed a significant lower number of anomalous points out of the 2-points range lines traced parallel to the trend one: 31% against 22% for 2018 and 2019, respectively. In fact, only three results can be classified as very anomalous in 2019 (highlighted in black), showing a clear discordance between CT and global score. For lab work and theory exam (not shown), the percentage are similar, with about 10% less out-of-range point for the second year.

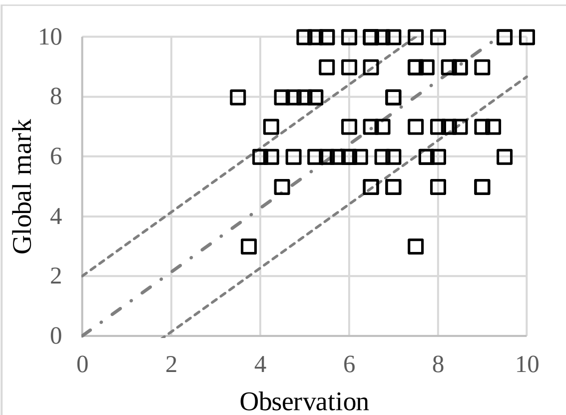

Figure 1 Observation instrument rates vs. global marks of the subject for the first year (2018)

https://doi.org/10.4995/muse.2022.17908
e-ISSN: 2341-2593

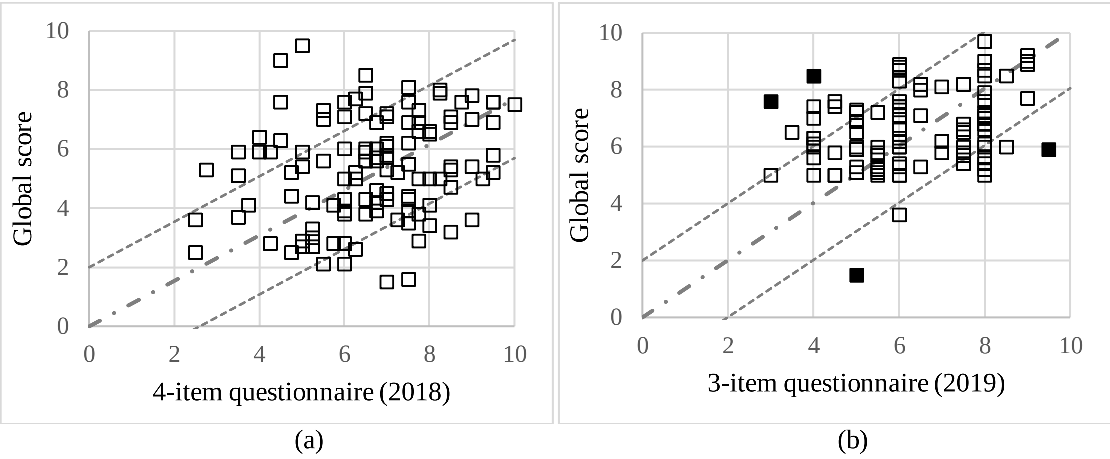

Figure 2 Questionnaire vs. global scores for the 4-item in 2018 (a) and 3-item (b) instruments conducted

In order to delve into this trend, we divided the performance scores into Failure (D': 0–4.9 points), Sufficient (C': 5–6.9), Good (B': 7–8.9) and Excellent (A': 9–10). The numerical values for the CT were obtained by scoring the 4-item questionnaire at 80% and observation at 20% for the first year; the 3-item questionnaire involves 100% of the CT grade. These values are classified within the A, B, C and D levels of achievement defined in the previous section. From these intervals, we considered an ideal correlation to get the pair A-A' for outstanding students, B-B' for the remarkable ones, C-C' for who barely passed the subject and D-D' for the failed ones. Distributing the students' scores within these intervals, we obtain **Figure 3** for both years experienced.

In the first year, it was surprising to find in **Table 4(a)** the only 2 students with A' qualification in the subject with C grade in CT, or 12 D' with A. A 76% of the students were located in the upper left corner that, although partially explained due to the lower mean of the global score average of more than 1 point, permits to visualize a poor correlation between both subject and CT scores. In contrast, we observe a significant shift in the distribution of the second year, more diagonal-based, with no-over 10 students in any box out of the main diagonal or sub-diagonals of **Table 4(b)**. Only 4 students with A' in the subject got A in CT, and there were 14 students in the B-B' box and 20 in the C-C' one. The jarring note was getting 3 students with A who did not pass the subject. It is also interesting to highlight that only a 9% got lower grades in CT than in the subject, suggesting that the restrictive selection criteria of students in Engineering degrees promote the access of students with good critical abilities.

https://doi.org/10.4995/muse.2022.17908
e-ISSN: 2341-2593

Table 4 CT vs. global performance distributed in the intervals defined previously for the: (a) first year (2018), (b) second year (2019) experienced

A	12	18	12	0
B	16	18	8	0
C	12	8	3	2
D	4	3	0	0
CT / subject	D'	C'	B'	A'

(a)

A	3	9	11	4
B	1	21	14	0
C	8	20	9	0
D	0	2	1	0
CT / subject	D'	C'	B'	A'

(b)

For the analysis of the instructor bias in the CT correction, it is proposed a correlation factor defined through the weight matrix W presented in **Table 5**. It weights with 1 the number of students located in the diagonal boxes (A-A', B-B', C-C', D-D'), and the farther the rest of the boxes from the diagonal is, the lesser the weight is (0.5, 0.25 and 0). By multiplying box-to-box the 4x4-matrix W with weights of each box (**Table 5**), and the 4x4-matrix p with the number of each box (**Table 4**), we get the factor:

$$\rho = \frac{\sum_{i=1}^{4}\sum_{j=1}^{4} W_{ij} p_{ij}}{\sum_{i=1}^{4}\sum_{j=1}^{4} p_{ij}} \quad (1)$$

which gives 1 if all the CT-subject points are contained in the diagonal boxes (perfect correlation using this estimative criterion) and 0 if they are in the A-D' and/or D-A' positions (the worst possible correlation).

Table 5 Weight matrix W to compute the CT vs. scores correlation from Eq. (1)

A	0	0.25	0.5	1
B	0.25	0.5	1	0.5
C	0.5	1	0.5	0.25
D	1	0.5	0.25	0
CT / Subject	D'	C'	B'	A'

https://doi.org/10.4995/muse.2022.17908
e-ISSN: 2341-2593

Table 6 gathers this correlation factor for the three performance indicators considered, in average and separated per instructor. First, it is confirmed that the correlations for the second year are significantly more consistent in line with **Figures 1** and **2** (excluding the specific case of Instructor 2 for the theory exam). The standard deviation is only 0.02 (2018) and 0.01 (2019) for the lab sessions, while this value is slightly more significant for the theory and global scores. Internally, these results indicate us that successive meetings and feedback between lecturers minimize the instructor bias, especially in the second year.

Table 6 Correlation ρ −Eq. (1)− between CT vs. lab work, theory and global scores for the two years (2018 and 2019) experienced, with the average and the standard deviation σ from the three instructors

Theory	**Instructor 1**	**Instructor 2**	**Instructor 3**	**avg.**	σ
1st year	0.41	0.47	0.33	**0.39**	0.07
2nd year	0.54	*0.46*	0.49	**0.51**	0.04
Lab work					
1st year	0.62	0.62	0.65	**0.63**	0.02
2nd year	0.68	0.67	0.69	**0.68**	0.01
Global					
1st year	0.52	0.49	0.40	**0.46**	0.06
2nd year	0.66	0.61	0.71	**0.65**	0.05

4. Discussion

4.1. Open questionnaire

The poor correlations given by **Figure 2(a)** for the open questionnaire in the first year indicated us that the design had to be reviewed. We concluded that the item 4 was not well outlined, since the students made a list of the parameters that influence the dynamic study without discerning the key elements or the plot line needed to address general dynamic cases. This generality of the question led to memory-based answers without reasoning implications. After deleting this question, the second

https://doi.org/10.4995/muse.2022.17908
e-ISSN: 2341-2593

year brought about better correlations (see **Figure 2(b)**), revealing that CT plays a significant role in the skills that the student needs to pass the subject, as we expected.

The influence of the instructors was also critical to explain these better results. There were three instructors in the lab sessions, who were responsible for promoting and assessing CT in their respective groups. The **Results** section analyze the bias per instructor and year (see **Table 6**), giving low values for both years, but better correlation for the second one. The coordination achieved in successive meetings to prepare the subject allowed the understanding required to minimize the bias.

By analyzing the results in detail, more important deviations are observed (poor correlations) in students with lower performance in the subject (see **Table 4**). These outcomes could suggest that some students with good CT have not prepared the subject properly to address the exams. Perhaps, they consider that the skills related to CT that they are aware of possessing will facilitate them to pass the exams without the need of a great effort in preparing the subject. On the contrary, other students, who are aware of their difficulties in reasoning skills, need to spend more time on the subject even though their learning is based more on memorization than on understanding the underlying physical concepts. These are hypotheses that should be contrasted with individual interviews to a representative sample of the student body.

4.2. Observation

As part of the intervention activities in the immersion approach followed, debates were promoted within two or three pairs in order to look for strategies to overcome the problems arisen during the lab session. But these were revealed insufficient to observe the individual way of thinking of each student. Although we tried to promote the participation of the most reluctant students to speak, many of them were masked in the arguments of the most participative ones. Furthermore, working in pairs often results in the reasoning of one of the components being opaque. And we must include the difficulty of scoring objectively with a score from 0 to 10 the competence, based on the quality of their arguments and not on how often they intervene in the debate. Sometimes, to separate these concepts is a challenging task, especially in the global debates at the end of the lab session. Those factors explain the really indiscernible correlations of the observation between the subject

https://doi.org/10.4995/muse.2022.17908
e-ISSN: 2341-2593

indicators in the first year (see **Figure 1**), which led us not to repeat this assessment instrument in the second year.

5. Conclusions

A practical intervention has been carried out for promoting and assessing CT in a technical subject for two years. An Immersion approach has been used to work on the skills and to encourage the dispositions of CT. Two assessment instruments have been developed to deal with the specific characteristics of the subject, with more than one hundred students every year, and to obtain some quantitative feedback in order to measure the validity and reliability of the intervention. The correlation between CT development and students' performance has been analyzed from a qualitative point of view.

The main conclusions that can be drawn from this work are:

- The need for an appropriate design of the assessment instruments for the specific evaluation of CT in technical subjects with a large group of students. Assessing a complex competence such as CT in a challenging subject such as *Mechanics* requires a careful trade-off. The instruments should be able to measure the cognitive skills and dispositions of the students related to CT, but they should be embedded in the contents of the subject to avoid using extra-time for this evaluation.

- A reliable way to confront a cognitive competence as CT in large groups in engineering courses is to implement an immersion approach in lab sessions with a limited number of students (up to 25). In this experience, computer lab works are designed to offer plenty of opportunities to put the CT skills and dispositions into play.

- The open questionnaire has revealed a promising instrument to assess CT, adapting the associated learning outcomes to the technical nature of the engineering courses. The observation tool needs a more individualised intervention that is not feasible for the size of the lab groups which we dealt with.

https://doi.org/10.4995/muse.2022.17908
e-ISSN: 2341-2593

- Developing good CT skills provides a better opportunity to pass the subject successfully, but they are not the only skills required. From a qualitative point of view, a good correlation has been observed between the CT grade and the students' performance for those with the best grades in the subject. However, a rather poor correlation has been obtained for those students with an average performance in the subject. Anyway, it is expected that the prolongation of the experience in successive courses will confirm that the evaluation tool detects improvements in the development of the competence and that the direct relationship with the performance in the subject could allow in future courses the evaluation of this competence directly from a correlation with the mark of the subject. Further research is needed to check the relevance of promoting CT skills in the students' achievements.

- An effective coordination of the instructors is essential to minimize the bias in the teaching and in the assessment of CT. An important effort should be done to agree a set of learning outcomes and how to grade them.

Because of the importance of CT for engineering students, further research on its promotion and assessment will continue in the next years for developing better strategies and instruments. A suitable design of any intervention should be done to ensure that it is aligned with the learning outcomes desired both for the technical and the CT skills. Reducing the influence of the instructors should be a priority when dealing with large groups of students.

Acknowledgements: The authors acknowledge the financial contribution by the Universitat Politècnica de València through the project PIME/2018/DPTO. IMM.

Author Contributions: Conceptualization, J. G and J. C; methodology, J. G, A. S, J. M-C., J. C.; validation, J. G, A. S, J. M-C., J. C. ;formal analysis, J. G, A. S, J. M-C., J. C.; investigation, J. G, A. S, J. M-C., J. C.; resources, J. C. ; data curation, J. G.; writing original draft preparation, J. G., A. S, J. M-C., J. C.; writing review and editing, J. G., A. S, J. M-C., J. C.; visualization, J. G., A. S, J. M-C., J. C.

Funding: This research was funded by the Universitat Politècnica de València through the project PIME/2018/DPTO. IMM.

Conflicts of Interest: "The authors declare no conflict of interest." "The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results".

https://doi.org/10.4995/muse.2022.17908
e-ISSN: 2341-2593

References

Ahern, A., Dominguez, C., McNally, C., O'Sullivan, J. J. and Pedrosa, D. (2019). A literature review of critical thinking in engineering education. Studies in Higher Education, 44, 816–828. https://doi.org/10.1080/03075079.2019.1586325

Ahern, A., O'Connor, T., Mac Ruairc, G., McNamara, M. and O'Donnell, D. (2012). Critical Thinking in the University Curriculum: The Impact on Engineering Education. European Journal of Engineering Education, 37 (2), 125–132. https://doi.org/10.1080/03043797.2012.666516

Andrews, J., and Higson, H. (2008). Graduate employability, 'Soft skills' versus 'Hard' business knowledge: A European study. Higher Education in Europe, 33, 411–422. https://doi.org/10.1080/03797720802522627

Angeli, C., and Valanides, N. (2009). Instructional Effects on Critical Thinking: Performance on Ill-Defined Issues. Learning and Instruction, 19 (4), 322–334. https://doi.org/10.1016/j.learninstruc.2008.06.010

Baytiyeh, H. and Naja, M. K. (2017). Students' Perceptions of the Flipped Classroom Model in an Engineering Course: A Case Study. European Journal of Engineering Education, 42 (6), 1048–1061. https://doi.org/10.1080/03043797.2016.1252905

Birenbaum, M. (1996). Assessment 2000: Towards a pluralistic approach to assessment. Alternatives in assessment of achievements, learning processes prior knowledge. M. Birenbaum and F. J. R. C. Dochy, eds., Kluwer, Boston, 3–30. https://doi.org/10.1007/978-94-011-0657-3_1

Dominguez, C. coord. (2018). A European Collection of the Critical Thinking Skills and Dispositions Needed in Different Professional Fields for the 21st Century. ITAD, Vila Real.

Douglas, K. A., Fernández, T., Purzer, S., Fosmire, M., and Van Epps, A. (2018). The Critical-Thinking Engineering Information Literacy Test (CELT): A Validation Study for Fair Use Among Diverse Students. International Journal of Engineering Education, 34 (4), 1347–1362.

Dunne, G. (2015). Beyond critical thinking to critical being: Criticality in higher education and life. International Journal of Education Research, 86–99. https://doi.org/10.1016/j.ijer.2015.03.003

Facione, P. A. (1990). Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction, executive summary, The Delphi Report. The complete American Philosophical Association Delphi research report, California Academic, Millbrae, CA.

Guerra, A., and Holgaard, J. E. (2016). Enhancing Critical Thinking in a PBL Environment. International Journal of Engineering Education, 32 (1(B), 424–437.

Hagerty, D. J., and Rockaway, T. D. (2012). Adapting entry-level engineering courses to emphasize critical thinking. International Journal of STEM Education, 13 (2), 25–34.

Hart Research Associates. (2013). It takes more than a major: Employer priorities for college learning and student success. Accessed February 10, 2021. http://www.aacu.org/liberaleducation/le-sp13/hartresearchassociates.cfm.

Hong, Y. C. and Choi, I. (2015). Assessing reflective thinking in solving design problems: The development of a questionnaire. British Journal Education Technology 46 (4): 848–863. https://doi.org/10.1111/bjet.12181

Itabashi-Campbell, R., Perelli, S., and Gluesing, J. (2011). Engineering Problem Solving and Knowledge Creation: An Epistemological Perspective. Technology Management Conference (ITMC), IEEE International, San Jose, California, 777–789. https://doi.org/10.1109/ITMC.2011.5996057

https://doi.org/10.4995/muse.2022.17908
e-ISSN: 2341-2593

Jackson, D. (2010). An international profile of industry-relevant competencies and skill gaps in modern graduates." International Journal of Management Education 8 (3): 29–58. https://doi.org/10.3794/ijme.83.288.

Jonassen, D., Strobel, J. and Beng Lee C. (2006). Everyday Problem Solving in Engineering: Lessons for Engineering Educators. Journal of Engineering Education 92 (2): 139–151. https://doi.org/10.1002/j.2168-9830.2006.tb00885.x

Kelly, A. (2001). The evolution of key skills: towards a tawney paradigm. Journal of Vocational Education & Training 53 (1): 21–36. https://doi.org/10.1080/13636820100200149

Liu, O. L., Frankel, L. and Roohr, K. C. (2014). Assessing Critical Thinking in Higher Education: Current State and Directions for Next-Generation Assessment. ETS Research Report Series 1: 1–23.

Lönngren, J., and Svanström, M. (2015). Systems Thinking for Dealing with Wicked Sustainability Problems: Beyond Functionalist Approaches. In Proc., 7th International Conference on Engineering Education for Sustainable Development (EESD'15) 1–7. University of British Columbia. https://doi.org/10.1007/978-3-319-32933-8_14

Llopis-Albert, C., Rubio, F., Zeng, Shouzhen, Z., Grima-Olmedo, J., Grima-Olmedo, C. (2022) The Sustainable Development Goals (SDGs) applied to Mechanical Engineering. Multidisciplinary Journal for Education Social and Technological Sciences 9(1), 59-70. https://doi.org/10.4995/muse.2022.17269

McCormick, M., Lawyer, K., Wiggins, J., Swan, C. W., Paterson, K. G. and Bielefeldt, A. R. (2015). Sustainable Engineering Assessment Using Rubric-Based Analysis of Challenge Question Responses. Journal of Professional Issues in Education and Practice 141 (2).

Michaluk, L. M., Martens, J., Damron, R. L. and High, K. A. (2016). Developing a Methodology for Teaching and Evaluating Critical Thinking Skills in First-Year Engineering Students. International Journal of Engineering Education 32 (1(A)), 84–99.

Naimpally, A., Ramachandran, H. and Smith, C. (2012). Lifelong Learning for Engineers and Scientists in the Information Age. Elsevier, Amsterdam. https://doi.org/10.1016/C2010-0-67009-X

Paul, R. (1995). Critical thinking: How to prepare students for a rapidly changing world, Foundation for Critical Thinking, Santa Rosa, CA: Ritthenhouse Book Distribu.

Pierce, C. E., Gassman, S. L., and Huffman, J. T. (2013). Environments for Fostering Effective Critical Thinking in Geotechnical Engineering Education (Geo-EFFECTs). European Journal of Engineering Education 38 (3), 281–299. https://doi.org/10.1080/03043797.2013.800021

Rieckmann, M. (2012). Future-oriented higher education: Which key competencies should be fostered through university teaching and learning? Futures 44: 127–135. https://doi.org/10.1016/j.futures.2011.09.005

Rodríguez-Largacha, M. J., García-Flores, F. M., Fernández-Sánchez, G., Fernández-Heredia, A., Millán, M. A., Martínez, J. M., Vilardaga, I. and Bernaldo, M. O. (2015). Improving Student Participation and Motivation in the Learning Process. Journal of Professional Issues in Education and Practice 141 (1). https://doi.org/10.1061/(ASCE)EI.1943-5541.0000209

Schulz, B. (2008). The importance of soft skills: Education beyond academic knowledge. Windhoek, Namibia: Dept. of Communication, Polytechnic of Namibia. https://doi.org/10.1016/0006-3207(93)90452-7

Sirotiak, T., and Sharma, A. (2019). Problem-Based Learning for Adaptability and Management Skills. Journal of Professional Issues in Education and Practice 145 (4). https://doi.org/10.1061/(ASCE)EI.1943-5541.0000420

Stassen, M. L. A., Herrington, A., and Henderson, L. (2011). Defining critical thinking in higher education. Chapter 10, To improve the academy, J. Miller, ed., Vol. 30, Jossey-Bass, San Francisco, CA, 126–141.

https://doi.org/10.4995/muse.2022.17908
e-ISSN: 2341-2593

Sursock, A., and Smidt, H. (2010). Trends 2010: A decade of change in European higher education. European University Association, Brussels.

UPV (2020). Transversal competencies. Accessed June 01, 2020. http://www.upv.es/contenidos/COMPTRAN/index-en.html

Young, K., and Chapman, E. (2010). Generic competency frameworks: A brief historical overview. Education Research and Perspectives 37 (1): 1–24.

Zlatkin-Troitschanskaia, O., Shavelson, R. J. and Kuhn, C. (2015). The international state of research on measurement of competency in higher education. Studies of Higher Education 40 (3): 393–411. https://doi.org/10.1080/03075079.2015.1004241