European Journal of Engineering Education

ISSN: 0304-3797 (Print) 1469-5898 (Online) Journal homepage: http://www.tandfonline.com/loi/ceee20

Adding value to the learning process by online peer review activities: towards the elaboration of a methodology to promote critical thinking in future engineers

Caroline Dominguez, Maria M. Nascimento, Rita Payan-Carreira, Gonçalo Cruz, Helena Silva, José Lopes, Maria da Felicidade A. Morais & Eva Morais

To cite this article: Caroline Dominguez, Maria M. Nascimento, Rita Payan-Carreira, Gonçalo Cruz, Helena Silva, José Lopes, Maria da Felicidade A. Morais & Eva Morais (2015) Adding value to the learning process by online peer review activities: towards the elaboration of a methodology to promote critical thinking in future engineers, European Journal of Engineering Education, 40:5, 573-591, DOI: 10.1080/03043797.2014.987649

To link to this article: http://dx.doi.org/10.1080/03043797.2014.987649

Published online: 08 Dec 2014.

European Journal of Engineering Education, 2015
Vol. 40, No. 5, 573–591, http://dx.doi.org/10.1080/03043797.2014.987649

Taylor & Francis
Taylor & Francis Group

Adding value to the learning process by online peer review activities: towards the elaboration of a methodology to promote critical thinking in future engineers

Caroline Dominguez[a,b], Maria M. Nascimento[a,b*], Rita Payan-Carreira[c], Gonçalo Cruz[a], Helena Silva[d], José Lopes[d], Maria da Felicidade A. Morais[d] and Eva Morais[a]

[a]*ECT, Universidade de Trás-os-Montes e Alto Douro, Vila Real, Portugal;* [b]*CIDTFF (Didatics and Technology in Education of Trainers, UA), Aveiro, Portugal;* [c]*ECAV, Universidade de Trás-os-Montes e Alto Douro, Vila Real, Portugal;* [d]*ECHS, Universidade de Trás-os-Montes e Alto Douro, Vila Real, Portugal*

(*Received 8 December 2013; accepted 13 October 2014*)

Considering the results of research on the benefits and difficulties of peer review, this paper describes how teaching faculty, interested in endorsing the acquisition of communication and critical thinking (CT) skills among engineering students, has been implementing a learning methodology throughout online peer review activities. While introducing a new methodology, it is important to weight the advantages found and the conditions that might have restrained the activity outcomes, thereby modulating its overall efficiency. Our results show that several factors are decisive for the success of the methodology: the use of specific and detailed orientation guidelines for CT skills, the students' training on how to deliver a meaningful feedback, the opportunity to counter-argument, the selection of good assignments' examples, and the constant teacher's monitoring of the activity. Results also tackle other aspects of the methodology such as the thinking skills evaluation tools (grades and tests) that most suit our reality. An improved methodology is proposed taking in account the encountered limitations, thus offering the possibility to other interested institutions to use/test and/or improve it.

Keywords: web-based peer review; peer feedback; critical thinking; engineering education; higher education

1. Introduction

The call for a change, either from national or from international organisations, about how engineers are educated, is well documented, and has evolved in different ways (Froyd, Wankat, and Smith 2012). According to the Accreditation Board for Engineering and Technology, the Association for Computing Machinery, and the Institute of Electrical and Electronics Engineers, among others, there is an emphasis to move towards learner-centred, cooperative, and technology-enhanced learning approaches, which the curricula should endorse, addressing active learning methods, promoting lifelong engineers with effective communication skills, both oral and written, with problem-solving, project-driving, and team-building capacities (Litzinger et al. 2011; Claris and Riley 2012). The failure to address such skills during graduation results in the current

*Corresponding author. Email: mmsn@utad.pt

existing gap in the transition to the labour market (Stein et al. 2007; Nair, Patil, and Mertova 2009).

When they reach university, as stated by Karandinou (2012), undergraduate students show limited critical thinking (CT) and communication skills. In this sense, writing based on reflection reveals itself as an effective way to promote them (Cooney, Alfrey, and Owens 2008; Ralston and Bays 2010), as well as peer review (Bauer et al. 2009; Sondergaard 2009) or the use of web-based collaborative environments (Brodahl, Hadjerrouit, and Hansen 2011; Calvo et al. 2011). This combination of methods offers students the opportunity to give and receive feedback from the colleagues, while they digest, analyse, and judge content, statements, inferences, and premises, arguing their thoughts in order to achieve a valuable conclusion and/or solution about a complex problem in an active, continuous, cooperative learning process.

Trailing previous studies with engineering classes (Dominguez et al. 2013; da Silva Nascimento 2014), this paper presents the final results of an online writing peer review methodology developed for engineering educators that contributes to enhance students CT skills.

The authors used an action research approach (Fahim and Nazari 2012) as it allows, in a dynamic and flexible way, to extract information, make small adjustments while the activity is progressing, and to tailor the experiment to particular situations (such as the number of students in class and the core subject in the graduation, among others). The constant monitoring of the activity throughout different cycles ensured the progressive inclusion of improvements from one cycle to another.

In this paper, a reflection is developed around the factors underlying the progressive modification of the learning adopted methodology, the elements that influenced the development of students' communication and CT skills in an online peer review environment, and finally around the limitations of CT evaluation tools.

Supported by an initial background section, the description of the methodology in use and the analysis of the results gathered until now are followed by the discussion of results and conclusions.

2. Background

The globalisation of the engineering profession associated to the growing of students' mobility, the economic crisis, and the prompt change of information and communication technology is creating a worldwide competitive market forcing the reconsidering of the role of future engineers. Engineers need to acquire more than technical skills, adding value to their enterprises with other professional skills, such as CT (Shuman, Besterfield-Sacre, and McGourty 2005).

CT is strongly related to complex problem-solving situations (Saiz and Rivas 2008), particularly in the engineering subjects. It is familiar to most educators, but not easily defined (Ralston and Bays 2010). According to the literature, CT is conceptualised based on multiple dispositions and skills (Angeli and Valanides 2009). It corresponds to a reflexive and proactive approach to any situation, going beyond simple intuition and perception, which in association with a critical judgement (including the interpretation of information and the quality of inferences and assumptions presented) will lead to different decision-making patterns, innovative recommendations, or the proposition of alternative solutions (Walker 2003). Thus, beyond a set of skills, and based on the Ennis definition (1997), we will consider CT as a 'reasonable, reflective thinking focused on deciding what to believe or do', stressing therefore the capacity of wise decision-making, either in the world of ideas or in the world of actions.

Instructional methods are essential to engage students in CT activities. Educators must consider the learning experiences and teaching approaches that promote better CT understanding.

There are clear disconnected relationships between what educators considered CT experience and what students identified as practice (Cooney, Alfrey, and Owens 2008). For instance, the need to clarify student learning desired outcomes and what will count as evidence and criteria highlights the value of teachers' support and monitoring in the process (Broadbear 2003). Additional support should be provided by explicit instruction with lectures designed specifically for guidance on the general CT skills, which may or may not include some practicing (Marin and Halpern 2011). This introductory approach is of utmost importance compared to unguided exploration, in particular when the situation in analysis is poorly defined or controversial (Angeli and Valanides 2009). CT may grow and be reinforced with time as a reflex of the stimulatory activities and the engagement of the participants. Also, it can fade if not nurtured, thus suggesting the importance of repeated practice (Marin and Halpern 2011).

Research also evidences that instruction supporting CT uses questioning techniques instead of lecture and rote memorisation (Kang and Howren 2004). Among such strategies, the FRISCO (named on the acronym for the identified aspects of critical reasoning: Focus, Reasons, Inference, Situation, Clarity, and Overview) (Ennis 1996) and the Paul-Elder guidelines (Paul and Elder 2003), or even the 'IDEALS' technique (Facione 2011) are oftentimes cited. In the case of writing based on reflection activities, educators should not only design writing assignments requiring students to demonstrate a synthesis of material, evaluate arguments, and deduce conclusions, but more important is the 're-writing' due to the direct inclusion of feedback and counter-argumentation (Tsui 2002). Since paired students show higher skills in writing and CT than unpaired (Ralston and Bays 2010), the peer review is seen as a tool for active learning (Knight and Steinbach 2011) and improvement of skills associated with diagnosis, evaluation, synthesis, and communication of CT (Bauer et al. 2009; Sondergaard 2009) while it also promotes the development of writing abilities (Ozogul and Sullivan 2009).

To avoid constraints, peer-review should be anonymous (Lu and Bol 2007). Nevertheless, other major factors are pointed out as strong modulators of the feedback quality, namely the students' personal considerations on their colleagues' competencies for feedback, the existing trust amongst peers (Hattie and Timperley 2007), the type of feedback, the students' perception of fairness in the process (Sung et al. 2005), the self-confidence (Walker 2003), and also the individual disposition to engage in the activity (Mwalongo 2011). Consequently, in order to support and enhance the feedback quality, the model proposed by Nelson and Schunn (N&S) reveals itself as a tool that can be presented and discussed with the students (Cruz et al. 2013). This model indicates the possible influence of the cognitive and affective features of a feedback on its understanding and agreement (Nelson and Schunn 2009).

However, the integration of CT in the curriculum through peer review activities is not an easy task for teachers and multiple barriers can hamper it (Fani 2011): the lack of teacher training and proper assessment of the activity (Ennis 2003), the lack of information and conceptualisation of CT (Scriven and Paul 2007), the preconceptions and attitudes towards CT (Kang and Howren 2004), and time constraints (Brodie and Irving 2007). Successful strategies to overtake these obstacles should be adopted, including the use of multiple-format assessment tests (Ku 2009), specific teacher training courses (Pithers and Soden 2000), and the consideration of different students' CT levels (Greenlaw and Stephen 2003), encouraging students' positive attitude (Browne and Freeman 2000).

Finally, a major discussion about CT concerns its assessment (Ku 2009). How can we measure the success of our efforts? How should we test CT abilities in students? For that purpose, literature refers to several tests of distinct formats (Arter and Salmon 1987; Carpenter and Doig 1988). Among all, some of the most popular are the Cornell Critical Thinking Test (CCTT; Ennis and Millman 1985), the California Critical Thinking Skills Test (CCTST; Facione 1990), and the Halpern Critical Thinking Assessment (HCTAES; Halpern 2010). The formers include a specific number of multiple-choice questions, based on general context situations or imaginary scenarios (respectively CCTST and CCTT) and aim to analyse different categories of skills (interpretation, analysis, evaluation, inference, and explanation, for CCTST; and induction, deduction, credibility, assumptions, semantics, definition, and prediction, for CCTT) (Ku 2009; Butler et al. 2012). The use of those tests is however controversial. Some studies have pointed to a failure in consistency, validity or reliability, inability to detect changes within specific disciplines (Saiz and Rivas, 2008; Ku 2009; Behar-Horenstein 2011), and the incapacity of these tests to actually monitor the type of reasoning at stake from the responses (Saiz and Rivas 2008).

In contrast, HCTAES combines multiple-choice and open-ended questions. In an attempt to avoid some of the criticism pointed out to the previously mentioned tests, its questions are based on authentic and believable situations. The open-ended questions aim to analyse the consciously thinking process and the selection of appropriate knowledge to create the answer, while the objective of multiple-choice questions is to test the capacity to remember a possible approach or solution. The skills represented in this test are verbal reasoning, argument analysis, and hypothesis testing, along with decision-making and problem-solving (Ku 2009; Butler et al. 2012).

However, if some authors tend to treat CT as a generic intellectual capacity manifesting itself in a broad way across disciplines (Norris and Ennis 1989), others argued that it needs to be assessed in the contexts in which it occurs (Cromwell 1992). For the successful acquisition of such skills, some authors defend the need for specific-domain knowledge to fully explore it, particularly in complex problem-solving situations (Angeli and Valanides 2009; Lai 2011). In contrast, general-content tests are applied to estimate the levels of individual CT, thereby serving as predictors, despite the controversies on its suitability for evaluation of well-reasoned judgements within a particular domain (Deal and Pittman 2009). Assessments should emphasise thinking rather than facts (Ennis 1993). Subjective tools such as essay questions and case studies require students to apply their knowledge to new situations, and are better indicators of understanding than objective true/false or standardised multiple-choice assessments. However, instructors can create multiple-choice questions that require CT. For example, asking students to identify the example that best applies a specific concept requires more CT and analysis than a question that asks students to identify the correct term for a given definition (Norris 1989). Although multiple-choice tests require more time and work to create than the open-ended CT assessments, they are also easier to grade (Ennis 1993). CT tests using a single multiple-choice format measures only recognition or level of knowledge, and do not adequately capture the dispositional characteristics of test-takers (Ku 2009). The multiple-choice response format does not reveal the underlying reasoning for choosing a particular answer, nor does it reflect the ability to think critically under unprompted situations (Saiz and Rivas 2008). Thus, both the multiple-choice and open-ended tests of CT have individual limitations (Ku 2009; Butler et al. 2012). The current trend is to combine the two response formats into one test. Multiple factors can render invalid multiple-choice credibility judgment in tests that accept only one answer as correct, including differences in students' degrees of CT sophistication, age, gender, social background, extra-critical-thinking empirical beliefs, assumptions made during test taking, and political and religious ideologies (Norris 1989; Deal and Pittman 2009). An improved methodology ought to provide means of lessening the effects of such factors, allowing responses in both multiple-choice and open-ended formats, making it possible to assess individuals' spontaneous application of thinking skills on top of their ability to recognise a correct response.

Assessment of CT skills outcomes, whatever the method used, may prove helpful to evaluate the effectiveness of the activity on the gain of competencies, to identify problematic issues demanding correction, or to analyse the results of the feedback provided during the activity or the levels of satisfaction of the participants (Deal and Pittman 2009).

3. The implementation of a methodology of peer review for future engineers

The implementation of the peer review methodology here presented evolved from an activity proposed by a professor teaching Industrial Management semi-annual credit units to higher education engineering students in the first graduation cycle of multiple engineering courses (Informatics, Mechanics, Civil, and Energy) at the University of Trás-os-Montes e Alto Douro (UTAD, Portugal). Besides the cognitive objectives related to the curriculum, the application of this methodology aimed to develop and improve students' communication and CT skills through online peer review and to measure their efficiency. The design of the methodology has progressed continuously from the first moment of its implementation, three years ago (first semester 2011/2012) to the last and current semester (first semester 2013/2014). At this point, four principal stages of the methodology are identified. The first two stages were described and analysed in previous publications (see, e.g. Dominguez et al. 2013; da Silva Nascimento 2014, for more details). Therefore, we shall only present herein data concerning the final two stages. The research questions that aroused interest in each stage, from the evidenced analysis, have guided the changes and decisions made until the final model now in use.

3.1. *General design of the methodology*

The general methodological design of the activity integrates several elements (described below) that interact and make possible its gradual improvement. Some elements are transversal to multiple strategies, as also mentioned in the analysis of research on CT development in engineering: problem-solving, ethical decision-making, conducting experiments, and assessing the social impact of technology (Ralston and Bays 2010), or even conceptual maps design (Jamison 2005).

The activity under analysis was carried out as one component of the syllabus of Industrial Management which allows students to get acquainted with the business world and the main management functions of an industrial company, in particular to train students in the use of SWOT (identification of Strengths, Weaknesses, Opportunities, Threats) economic strategic analysis approach (Bressy and Konkuyt 2008).

3.1.1. *The activity: tasks to be performed*

This activity integrates three cycles per semester, each four weeks long. Each cycle includes: (1) the reading of an economic article from a credible source (e.g. an economy newspaper or magazine) dealing with recent news of an economic phenomenon (like the closure of a company or the internationalisation of an economic sector); (2) the production of a written document containing the summary, the analysis of the opportunities and threats for economic agents, the critical analysis with the issuing of a personal opinion (task for the student-author); (3) a peer-review assessment of that document by a student-reviewer, according to the N&S model; (4) the counter-argumentation stage in which the student-author has the opportunity to read and comment on the reviewer's feedback, improving his/her work on a voluntary basis and (5) the evaluation and classification of the work by the teacher.

3.1.2. *Adoption of a communication system and support*

Google Drive (GDrive) was the web-based environment chosen to support the learning activity due to its easy access, similarity, and ease of integration with the Microsoft Word environment (Rienzo and Han 2009). Besides avoiding the high transfer of papers between students, due to the number of interactions, this online system allows absent students to do their work on a deferred basis until the deadline. In addition, it registers the work and comments made by students in a continuous way, saving histories of the different versions of the document. It also allows students to make synchronous and instantaneous editions.

The GDrive environment is also used to share support documents (e.g. guidelines containing a description of the objectives/tasks to be performed, coordination grids, schedules of the different cycles, and examples on feedback quality) as well as to write the assignment and make the review on a unique shared template. The activity is preceded by an introductory face-to-face session of the online environment in order to get the students acquainted with it.

A template of the document to be produced serves as a work basis. It contains the three parts of the assignment, in a standardised and simple structure: (1) summary, (2) analysis of the economic variables and of threats and opportunities, and (3) critical opinion. It is shared between paired students and the teacher.

3.1.3. *Teacher's monitoring of the activity*

The analysis of the shared documents and the clarification of questions about the GDrive environment are regularly monitored during the cycle. General guidance is also transmitted orally on how to make a good feedback using the N&S model, focusing especially on the need of the peer comments to be encouraging and constructive.

3.1.4. *Evaluation of students' assignments*

All assignments are graded. For that purpose, documents must evidence the following students' competencies: (1) For authors, a summary with clear ideas and without repetitions, a complete identification of variables at stake, a presentation of the opportunities and threats, and also a well-formulated substantiated personal critique of the article. (2) On the other hand, reviewers must show that they checked all parts of the student-author's work, and made a review according to the N&S model.

3.1.5. *Instruments for the assessment of the activity*

It is important to analyse the impact of the activity, whether from an overall perspective or specifically focusing on the students' communication and CT skills. In a first moment, this analysis was achieved by means of two instruments: a survey applied to the students at the end of the semester and a grid with all of the students' feedback evidences in the assignments, based on the N&S model.

For our investigation, we elaborated our own survey after getting acquainted with other questionnaires of this type in the literature, for example from Xiao and Lucking (2008). It allowed the characterisation of the student profile (e.g. gender, age, and their use of web-based tools, including GDrive), and covered the students' perceptions on the adopted pedagogical approach (including the execution of tasks, ease of use of the online environment, time and workload, and usefulness of the support materials), as well as on the quality of the received feedback (e.g. its value and contribution to improve the quality of their writing and skills' acquisition). Finally, it also evaluated the overall satisfaction of the students with the activity. Some open questions identified the reasons for students having (or not) used the suggestions of their peers and difficulties, allowing us to deepen the analysis of the activity. Following the methodology of qualitative content analysis (Krippendorf 2013), all answers were read thoroughly in order to infer the categories of analysis. Then all the contents were analysed and assigned to the inferred categories. This process was done manually (no software was used) by two panels of judges with no significant differences in the results.

In parallel, the quality of students' communication skills was evaluated through the analysis of the feedback, namely the provision of a global vision, the identification of local and global problems, and solutions, the use of explanations and positive language, the counter-argumentation, and the implementation of suggestions. This analysis was performed by evaluating the items present on each work according to the categories used by the N&S model.

Targeting the assessment of the evolution of the level of CT abilities, in the final stages we included in the methodology the Level X of the CCTT (Ennis and Millman 1985) validated to the Portuguese reality (Oliveira 1992; Tenreiro-Vieira 2004). It was applied at the beginning and at the end of the activity. Comparison of the results of paired tests and of the grades for each cycle was performed using paired *t*-tests for the mean grades.

3.2. *Implementation and assessment of the methodology in Stages 3 and 4*

Previous expertise, issued from the two primordial stages of the methodology (Stages 1 and 2) and analysed in previous publications (see, e.g. Dominguez et al. 2013; da Silva Nascimento 2014, for more details), showed that: (1) students felt less self-confident as reviewers; (2) the perception on feedback differed between authors and reviewers' students; (3) the use of feedback by authors was influenced by the existence of intermediate grading in the cycle; (4) although students possessed satisfactory levels of written communication skills, reinforcement was needed particularly in areas like the identification of global problems and solutions or guidance on meaningful feedback, e.g. by demonstrating successful examples and defining the characteristics of a good feedback; (5) the lack of anonymity was an element of discomfort for the students; (6) the need to reduce the teacher workload; and (7) the need to strengthen in class the constituents of specific CT skills (capacity of synthesis, identification of reasons, inferences, and conclusions).

Although for most of the students this approach was completely new, it was evaluated with general satisfaction and no specific problems were stated concerning the used web environment. Based on this knowledge, the methodological approach entered in its third and fourth stages, with adjustments to the general design described above.

3.2.1. *Stage 3 (first semester of 2012/2013)*

Twenty-eight undergraduate Civil Engineering students were enrolled in Stage 3. New aspects were introduced in this experimental stage: (1) the writing and peer-review processes became anonymous; (2) the paper in analysis was proposed by the teacher and was identical for all the students (when in previous stages it was selected by the students on an individual basis), which allowed also levelling the teacher's work and comparing in a more accurate way the effort required to do the assignments; (3) inclusion of the CCTT, applying it at the beginning and at the end of the activity and (4) the FRISCO guidelines (Ennis and Goldman 1991) were included and explained to the students to support the writing, the emission of more founded critical opinions, and counter-argumentation (Figure 1). All other instruments of analysis (the survey, the feedback analysis according to the N&S model, and the grading of students work) were maintained. Students who did not submit their works on schedule were graded under 10 points (in a 20-points scale). The entire activity was developed in the GDrive, with the technical support of the UTAD e-learning team in maintaining anonymity.

The teacher's monitoring role remained practically unchanged (Dominguez et al. 2013). A pre-scheduling of deadlines for submission of papers by authors and reviewers was introduced to guide the activity. Examples of the best previous works were shown and discussed in the

Stages (school year, Semester)	First year of the bachelor	Number of students	Number of cycles of the activity	Type of revision	Template availability	Documents that support students' work	Class of technology support	Classes on critical thinking	Students' assessment	Cornell test, level X
Stage 3 (2012/13, 1st)	Civil engineering	28	3, all with the *Google Drive* and the same articles selected by the teacher	Blind with teacher' and e-learning team overload since *Google Drive* is not prepared for blind review	Yes	Yes, Nelson and Shunn model to give feedback and FRISCO guidelines to support critics	Yes, by the teacher	No, the teacher only gave general and oral guidance referring the available documents	Intermediate peer review by the students; final revision and grades by the teacher	Pre and post-test done (before and after the activity) in paper
Stage 4 (2012/13, 2nd and 2013/14, 1st)	Energy and Mechanics engineering Informatics engineering	18 27 46		Blind using pseudonyms e-mails invented by the students to guarantee anonymity	Yes, improved with links to the support documents			Yes, 2 classes of one and a half hour fostered by the teacher	Intermediate peer review by the students, with the possibility to counterargu-ment by either authors or reviewers; final revision and grades by the teacher	Pre (before the activity) in paper (after the activity) in the *Google Forms* on-line Pre and post-test done (before and after the activity) in the *Google Forms* on-line

Figure 1. Elements of the methodology at each stage.

classroom although not specifically trained before the beginning of the activity in a hands-on way.

The success of the activity was assessed using the four main instruments available during the overall activity: the survey, the analysis of feedback, the teacher's grading of the students work, and the pre- and post-activity Cornell test.

Out of the 28 students enrolled, 24 (86%) responded to the final survey. Once more, they showed a general satisfaction (71%) in relation to the overall activity. The majority (81%) had not used GDrive before and found the technology class useful/very useful (81%). As in other stages, students liked to be assessed by their colleagues (71%) but only 10% thought that colleagues assessed as well as the teacher. Students expressed however some difficulties in understanding the FRISCO guidelines, especially the 'inferences' (33%) and 'overview' (23%) elements, suggesting the need to strengthen the support in those points; 86% of the students agreed that they should continue this type of activities in other courses. Moreover, anonymity was appreciated by the students, by reducing the background bias associated with social and personal relationships. However, the form found for anonymity in the Drive (duplication of assignments) increased the workload for the teacher and her dependence on the support of the e-learning team.

Nine randomly selected students, corresponding to 37.5% of surveyed students, were used to deepen the analysis of the responses with the qualitative questions of the survey. The main reasons for students' positive perceptions on being evaluated by their peers were the improvement of their learning process – e.g. 'because when we are evaluated by a colleague and not by the teacher, it helps us having another perspective on the work' (Student 5) or 'it is a different way to be assessed and since we are from similar ages, we have similar forms of thinking' (Student 7). On another hand, some felt unfairness feelings resulting from the peer evaluation, and therefore did not like to be evaluated by their peers.

As reviewers, students liked to assess their colleagues work – e.g. 'because I had to perceive what his ideas were and compare them with mine' (Student 1) or 'because since the student-author was not specific and was even irresponsible, I had to make an extra effort to analyse everything and it ended up to be good for me' (Student 2). Two students said that they would accept this methodology in future academic works because 'it is a simple and an innovative way to do some works and to know the opinion of other colleagues'. The characteristics of the activity seem to be a positive factor for students' satisfaction: e.g. 'the activity was well conceived and

Classes		Total (sum of all)	Total number of occurrences		
			Summary	Analysis	FRISCO
Summarization		3	0	0	3
Problems	Global	0	0	0	0
	Local	46	13	31	2
Solutions	Global	0	0	0	0
	Local	38	11	25	2
Explanations		38	8	17	13
Praises		7	3	0	4
Mitigation		8	5	1	2
Contra-argument		23	7	9	7
Feedback use		23	7	9	7

Figure 2. Total number of occurrences per categories of Nelson and Schunn in Stage 3 (nine assignments analysed).

forces us to analyse texts, expanding our capacity to interpret them better. If my opinion can help others: I am very open to collaborate' (Student 2).

The survey also showed that the main reasons for student-authors to use the peer suggestions concerned the feeling of the improvement of their work and to the quality of the feedback: 'the opinion of my colleagues allowed me to review my own opinion and see the errors I had made; although I thought I was right at first' (Student 5). To the question 'do you think that FRISCO guidelines were useful for your thinking process and in developing the critique?', students answered 'it helped developing CT skills in an efficient way' (Student 7) or served 'as a guide for their thinking process' (Student 1). Yet, some wrote that available information on the FRISCO guidelines 'was not clear' (Student 2). The main problems highlighted by students to do the activity were the 'lack of time' or 'having missed the first sessions (where the objectives and the methodology were presented)' (Student 2). They would also suggest to 'vary the type of papers chosen by the teacher so as to be more motivating (Students 6 and 7)', 'to get more clarification from the teacher on the FRISCO guidelines' (Student 2), and 'to compel students to do all parts of the work so as to make sense of the whole assignment' (Student 1).

The analysis of feedback (Figure 2) was performed in nine assignments (11% of all assignments), showing that at the cognitive level, the students tend to focus their attention on more detailed features of the N&S model, such as the identification of local problems and the proposal for their resolution, at the expense of the tasks of summarisation or of global problems' identification. The feedback given by the student-reviewer assumes forms of high specificity, illustrated by the differences between global and specific records of observations (Figure 2). At the affective level, the low number of occurrences of praise and mitigation suggests little concern from student-reviewers with aspects of linguistic politeness and courtesy. We can note however the existence of a large number of sequences with explanatory nature, either for justification of the proposed solutions, or as reinforcement of the student-reviewer agreement with the work of the student-author. The interaction between peers is noteworthy (more than in the initial stages of the methodology (Dominguez et al. 2013), increasing the counter-argumentation and the use of feedback occurrences. This can be attributed to the improvement of the template that became more structured, with specific spaces and guidelines for the application of the FRISCO critical analysis.

The teacher's final grading of the assignments was analysed and showed slight improvement between cycles (Figure 3). For this analysis, only the positive grades were used, censoring data from students having negative grades or not having accomplished all the tasks. Data showed that the difference between the mean grades of the first papers (MP1) and the third papers (MP3) was not significant (p-value $> 5\%$) despite that the mean third assignments' grades were slightly higher than those of the first (the blue negative value of the paired t statistic, Figure 3). In contrast,

Information	Civil (2012/13)		
	Paper 1	Paper 2	Paper 3
Mean (M)	14,39	13,28	15,27
Standard deviation	2,12	2,14	1,87
n	22		
Paired - t	MP1-MP2	MP1-MP3	**MP2-MP3**
	2,00	-1,66	**-4,025**
p - value	2,9%	5,6%	**0,03%**

Figure 3. Analysis of the students' positive grades (cycles in Stage 3).

Stage	Bachelor	Year (Semester)	Mean paired values Pre-test – Post-test		
			n	Student t	p - value
3	Civil	2012/2013 (1st)	21	-2,54	0,97%
4	Energy	2012/2013 (2nd)	12	3,00	0,61%
	Mechanics	2012/2013 (2nd)	17	2,73	0,73%
	Informatics	2013/2014 (1st)	32	2,50	0,91%

Figure 4. Cornell pre- and post-tests for all courses (Stages 3 and 4).

the difference between the MP2 and MP3 was significant (bold blue *p*-value $< 1\%$), the mean grades for MP3 being higher than for MP2 (the negative value of the paired *t* statistic). Admitting that grades reveal the level of accomplishment of the tasks and of students' evolution, we might say that in this case they reveal an improvement from the second to the third assignment.

These results could be explained either by the fact that the teacher slightly over-rated the first assignments in order to foster motivation in students for this new activity or to the subjectivity factor inherent to the assessment (Linn and Gronlund 2000).

The results from 21-paired Cornell tests (Figure 4) were analysed (75% of the students enrolled in the activity) and the results show that the mean scores of the pre-test were lower than those of the post-tests (*p*-value $< 1\%$).

Taking all assessment data together, it may be concluded that there was a general increase in CT skills (Figures 3 and 4) due to the introduction of FRISCO guidelines and their use by the students, even though the analysis of the survey reported some difficulties in its use, in particular for Inferences and Overview.

At the end of this stage, the analysis of the results raised two questions: a technological one on how to find a simpler way to share documents anonymously in the Drive environment, since the anonymous way of working was to be maintained but ideally with reduced workload; and a pedagogical one on how to improve the students' understanding of the FRISCO guidelines and of CT in general. Moreover, it was important to test again the validity of the Cornell test to evaluate CT skills acquisition with more applications.

3.2.2. *Stage 4 (second semester of 2012/2013, first semester of 2013/2014)*

This stage corresponds to the implementation of the final methodology for this CT activity in three undergraduate course classes totalling 91 engineering students: energy ($n = 18$); mechanics ($n = 27$); and informatics ($n = 46$).

Analysing previous results, it was decided, at this stage, to enhance guidance on the use of FRISCO and CT guidelines. For that, two specific sessions of 90 minutes were introduced at the onset of the activity, presented by the teacher and worked within the class, focusing on the arguments, counter-argumentation, on CT standards, and especially the FRISCO guidelines. The template created for this activity was also improved with the introduction of hyperlinks to supporting documents and scheduling; while all the other instruments of the activity were maintained. Following the support to feedback strengthening, examples of good review and counter-argumentation presented in the class were also made available online. To overcome the workload associated with the anonymity, students created electronic mails using unidentifiable pseudonyms. This measure greatly simplified and streamlined the management of shared documents in GDrive. Finally, a few new questions were added to the final survey concerning the students' perceptions of the FRISCO guidelines and the Cornell test.

The analysis of the final survey was performed on the responses of 78 surveyed students (17 students from energy, 26 from mechanics, and 35 from informatics), corresponding to 86% of the students, failing to evidence significant differences between the bachelors. As reported earlier (see Dominguez et al. 2013; da Silva Nascimento 2014, about 2011/2012 civil and energy students), the general perceptions on the use of the GDrive environment were very satisfactory. In total, 97% of the students considered the activity important for their learning process, 85% found that it developed their critical reflection, and 82% agreed/fully agreed that it improved their synthesis skills. Also, 87% thought that it improved their respect for their colleagues' opinion and developed their sense of responsibility (72%), collaboration (76%), and use of technology (64%). The objectives and support given in class as well as the supporting written documents were considered important and enough. In total, 96% of the students attended the specific class on CT and found it useful/very useful. From those, 86% managed to understand better the sense of CT, 50% agreed/fully agreed that they understood the FRISCO guidelines, totally or partially, while 86% agreed/totally agreed to recognise that these guidelines, available in a linked document of the template, allowed them to improve their work.

From the survey, no significant differences were found between student-authors' and student-reviewers' perceptions. Authors perceived the feedback received as positive, motivational, corrective (confronting ideas, arguments, and identifying problems related to writing and to content), clear, fair, valid, trustworthy, reliable, and useful. It also showed that 91% of the student-authors used the feedback given by their peers and found the peer review as effective as the teacher's. Concerning the FRISCO guidelines use, 94% of the student-authors knew that it was mandatory, 98% did know the meaning of each acronym letter, either by attending the class (16%) or by accessing (81%) the document provided by the teacher. However, 42% of the student-authors still had difficulties in using it, mainly with the 'I' of inferences, 64%, or to the other letters: 'F' (5%), 'S' (5%), 'C' (8%), and 'O' (4%). In general, 94% of the students expressed that the FRISCO guidelines were useful/very useful to support their thinking and to develop their critical analysis; 81% said that the activity improved their CT skills. As reviewers, only a minority disliked reviewing their mates' work (7%) and a majority recognised that they had skills to provide feedback. Additionally, 93% of the students (whether authors or reviewers) felt confident in their roles. From the reviewers' perspective, they gave mainly a corrective feedback (confronting ideas, arguments, and identifying problems related with the writing or the content). The majority (70%) followed up the authors work and verified the authors' changes (72%). Only 53% of the reviewers fully used the FRISCO guidelines in their review and had problems with its use (58%). Among the encountered problems, 55% declared difficulties with the 'I' of inferences and with the other acronyms: 'F' (7%), 'R' (7%), 'S' (10%), 'C' (5%), and 'O' (16%). In total, 95% of the students expressed that the FRISCO guidelines were useful/very useful to develop and guide their commenting skills to the student-author's work. Also, 97% of them said that the activity improved/improved a lot their CT skills.

The qualitative analysis of 46% (36 students randomly chosen) of the respondents to some specific questions completes the global analysis of the survey at this stage. To the question of 'why did you use the feedback given by your colleagues?': 32 in 36 (89%) stated that they considered different opinions improved their work ('because the opinion of the others allows us to improve our work', Student 16), two students referred improvement of their skills and one motivation ('because it gives me more will to do better and better and to argument better', Student 13; and 'it creates an extra motivation', Student 17).

To the question of 'why do you think that the FRISCO guidelines were useful to support the CT process?' almost all the authors (34 in 37, 92%, stated that it promoted the acquisition of CT skills, e.g. 'because it helped me to structure the ideas showing me if the critics were well constructed and completed', Student 8). Similarly, almost all the reviewers who used the FRISCO guidelines found that they allowed them to organise their ideas and thinking, e.g. 'because it separates complex tasks in easier ones' (Student 12).

To the question 'what do you think about the Cornell test?' (which was only asked to the Informatics Engineering students), the answers were not conclusive. On one hand, 17 of all 35 surveyed students have a positive feeling towards the test ('I found it interesting because it challenged my ability to solve problems in specific situations', Student 13). On the other hand, 15 in 35 students showed dissatisfaction in relation to the test, either because of its length and confusing nature or because it was perceived as an uselessness tool to measure CT skills ('I still don't know how this test can evaluate our improvements', Student 6).

The analysis of feedback plus the counter-argumentation using the N&S model, along with the use of feedback occurrences between authors and reviewers (Figures 5 and 6) was performed on 38 randomly chosen assignments (14% of all the assignments). It highlighted an increase in the identification of global problems and the proposal of solutions, showing more skills of synthesis and analysis in comparison to the previous stage. The identification of global problems still occurred more often on the summary task. However, students continue to put more emphasis on identifying local problems but advance proposals for solving most of them. The concern to explain both the identified problems and the proposals for solutions in the analysis and the FRISCO area of the template was also evident, especially for local problems (142 occurrences), or even to express the accordance with the point of view of the student authors. In 114 possibilities of feedback and counter-argumentation (at least one possibility in each of the three parts of the assignment multiplied by the number of assignments), there were 98 counter-arguments and 99 uses of feedback (Figure 6), which is a quite high number of occurrences, showing clearly that the interactions between peers increased compared to the previous stage, and also that students considered the suggestions given by peers (either agreeing or disagreeing). Finally, students used positive affective feedback (even when disagreements occurred), which positively contributed to the whole work outcome. Some of the students' testimonies revealing the effects of the feedback and of the counter-argumentations are given:

> I'm glad I helped the author to understand the paper, but I think that in the part of the technological variables, the author didn't understand exactly my idea (student 11); 'Good analysis of the variables in question; the student reviewer addressed ideas that initially I didn't take as important. I agree with what the reviewer said (Student 25)

and

> Again, I agreed with the author, even when my ideas were rejected. The rejections were well explained. I must admit I was wrong. So here I want to give a word of appreciation to the author who shared a wonderful job of a highly structured and committed analysis of paper. (Student 16)

Classes (Nelson & Schunn, 2009)			Total (47 files)	Number of occurrences per bachelor			
				Civil (9 files)	Mechanics (9 files)	Energy (11 files)	Informatics (18 files)
Summary	Summarization		22	0	5	7	10
	Problems	Global	17	0	4	5	8
		Local	54	13	16	8	17
	Solutions	Global	9	0	3	3	3
		Local	41	11	12	6	12
	Explanations		41	8	13	11	9
	Praises		21	3	5	8	5
	Mitigation		10	5	2	2	1
	Contra-argument		39	7	7	11	14
	Feedback use		37	7	5	9	16
Analysis	Summarization		14	0	5	9	0
	Problems	Global	2	0	0	1	1
		Local	157	31	42	40	44
	Solutions	Global	0	0	0	0	0
		Local	132	25	36	32	39
	Explanations		93	17	25	39	12
	Praises		16	0	0	9	7
	Mitigation		12	1	2	4	5
	Contra-argument		42	9	9	10	14
	Feedback use		43	9	7	11	16
FRISCO	Summarization		19	3	5	9	2
	Problems	Global	6	0	0	5	1
		Local	69	2	17	16	34
	Solutions	Global	5	0	0	4	1
		Local	58	2	13	12	31
	Explanations		116	13	22	32	49
	Praises		36	4	0	17	15
	Mitigation		19	2	6	9	2
	Contra-argument		40	7	9	9	15
	Feedback use		42	7	8	11	16

Figure 5. Nelson and Schunn categories for Stages 3 and 4.

Classes		Total (sum of all)	Total number of occurrences		
			Summary	Analysis	FRISCO
Summarization		52	22	14	16
Problems	Global	25	17	2	6
	Local	234	41	126	67
Solutions	Global	14	9	0	5
	Local	193	30	107	56
Explanations		212	33	76	103
Praises		66	18	16	32
Mitigation		33	5	11	17
Contra-argument		98	32	33	33
Feedback use		99	30	34	35

Figure 6. Total number of occurrences per categories of Nelson and Schunn in Stage 4 (38 assignments analysed).

Analysing the final assignments grades given by the teacher (Figure 7), it was shown that only students from informatics fairly increased their scores between MP1 and MP2 and between MP1 and MP3. Although not significant (p-value $> 5\%$), only a very slight improvement in the mean

Information	Energy (2012/13)		
	Paper 1	Paper 2	Paper 3
Mean (M)	15,77	15,31	15,19
Standard deviation	1,51	2,04	1,28
n	13		
Paired - t	MP1-MP2	MP1-MP3	MP2-MP3
	0,993	1,778	0,250
p - value	17,0%	5,0%	40,4%
Information	Mechanics (2012/13)		
	Paper 1	Paper 2	Paper 3
Mean (M)	15,19	14,56	14,10
Standard deviation	2,20	1,58	1,90
n	16		
Paired - t	MP1-MP2	MP1-MP3	MP2-MP3
	0,993	1,130	1,778
p - value	17,0%	13,9%	5,0%
Information	Informatics (2013/14)		
	Paper 1	Paper 2	Paper 3
Mean (M)	13,67	14,30	13,89
Standard deviation	1,73	1,70	1,89
n	28		
Paired - t	MP1-MP2	MP1-MP3	MP2-MP3
	-1,403	-0,508	1,015
p - value	8,6%	30,8%	16,0%

Figure 7. Papers grades for each cycle and courses at Stage 4.

grades may explain the (blue) negative values in the paired *t* statistic. It is possible that these results reflect different levels of students' engagement in the activities, or in contrast it might be explained by differences in the students' workload, individual background, or to the fact that the last cycle was coincident with the global assessment period for the graduation.

The analysis of 57 paired Cornell pre- and post-tests (Figure 4, above) highlighted a more favourable pre-tests mean score, both in general and individually for each course showing that students had a significantly better pre-test mean score than the post-test mean score (p-values $< 1\%$) and that the Cornell test does not show any significant improvements. Students' complaints over the Cornell test in the final survey suggest a saturation for the length of the test. This led us to question ourselves on the consistency of the adopted methodology, in particular the Cornell test, which we will discuss in the next section.

4. Discussion, conclusions, and future work

The labour market and the society challenges Engineers to understand and act on increasingly complex different systems (technology, environment, and society) and demand for continuous development and mastering new skills (Nair, Patil, and Mertova 2009). Communication and CT skills provide students not only to leverage their learning in an academic environment but also in their future workplaces. When CT is substantially worked on in a rigorous way, it spurs new domains and guides professional reasoning through complex engineering questions and issues, whether technological, commercial, environmental, ethical, or social (Claris and Riley 2012).

Nevertheless, students do not naturally think/use CT tools and are not acquainted with the important questions they should ask themselves, teachers, colleagues, customers, or vendors, to deepen their understanding and refine their thinking process. That is why engineering curricula ought to include the training of these skills.

The work presented here reflects the evolution of a methodology intended to foster and enhance communication and CT skills in freshmen engineering students, aiming to contribute to the body of research on CT activities and specifically to get information on how web-peer review activities, while an active method of teaching/learning, can promote CT skills.

The implementation and analysis of the methodology described above allowed drawing some conclusions and opening up to future investigations. Compared to its primordial stages, the methodology presented herein gradually included new elements/tools, like the FRISCO guidelines to support the critical opinion of the students, a specific class on CT, as well as the presentation of good examples of assignments in the class, which were perceived very positively by the students. As pointed out in the literature, specific support, guidance, and explicit instructions are of outmost importance for the success of this type of activities (Sgro and Freeman 2008) and for the acquisition of the CT skills (Marin and Halpern 2011). The design of the activity including several rounds of feedback, argumentation, and counter-argumentation was mainly appreciated by the students who felt that their CT skills had increased in various domains (as synthesis, evaluation, relating reasons and conclusions) confirming that the inclusion of feedback and the rewriting is a powerful tool for CT development (Tsui 2002).

Although the GDrive environment was new for most students, digital competencies were not an obstacle for the peer-review activities, and students felt well supported by the teacher during the activity. Students' previous acquaintance with digital instruments certainly was a favourable condition (Brodahl, Hadjerrouit, and Hansen 2011). The solution to turn the activity anonymous in a simple way using pseudonyms email addresses ensured that social or inter-personal aspects would not interfere in the whole process. The use of GDrive allowed all the tasks to be performed and proved to be satisfactory, reinforcing conclusions of other studies (Bauer et al. 2009)

Preliminary considerations on the use of the N&S categories for a 'meaningful feedback', along with its constant monitoring were very important and contributed to the success of the activity and the growth and development of skills, particularly in communication. To be effective, feedback cannot be vague or too complex; otherwise it does not achieve its purposes and discourages those involved in the process (Lu and Bol 2007). It should be clear, objective, and provide specific guidelines and suggestions related to the goals; otherwise it loses efficiency. Moreover, it determines how the comments are received and if they are introduced (or not) into the document being drafted (Strijbos, Narciss, and Dünnebier 2010). The analysis of the assignments through the N&S model confirmed that students increasingly used the categories of 'good feedback' fostering the use of the same by their mates. Nevertheless, in accordance with Walker (2003) our findings stress the need to provide more support on the 'feedback giving aspect'.

An oral class demonstration with adequate examples from previous works available online worked as an additional incentive for students involved in the activity and pointed to desirable paths to explore in the following stages of this activity. Nevertheless, the analysis of the assignments and of the final survey (both quantitative and qualitative questions) stressed the need to deepen some CT tools with the students, namely the inferences, a major difficulty reported by the students. As stated in the literature (Fani 2011), it is not easy to integrate CT in the curriculum, and it is also our common belief that it is important for the teacher to gain more experience and proper training in particular areas (Pithers and Soden 2000). Teacher's training may cover the available instructional strategies and feedback guidance, including on how to deliver a meaningful feedback (teachers are more used in direct numerical grading) or on the concept of inferences, dimension where students showed more difficulties.

The limitations found in previous stages concerning the use of the classical numeric grading form and the N&S model in the assessment of this activity drove the introduction of a specific and expedite tool for the assessment of CT skills outcomes − the Cornell text (level X) − the only existing critical test validated so far for Portugal. This was associated with the use of FRISCO guidelines (from the same author, assuring the process coherence), which was endorsed by students as very important for improving performance and helpful in structuring the thinking process. Beyond students' opinion, data analysis of the performance tools show a tendency of CT skills improvement associated with the methodology presented herein.

The data analysis of the Cornell test raises a question on its sensitivity to the short-term development of the CT skills. Indeed only slight improvements of the global performances could be reported by the test. This may be explained by the motivation of the students to engage in this test, which as revealed in the analysis, was considered by half of the students as too long and confusing. As discussed in literature (Ku 2009; Butler et al. 2012), another reason may be related to test format (multiple-choice) which, although easier to grade (Ennis 1993), might not show test takers' reasoning for choosing a particular answer. Additionally, the type of questions made, either too specific or too far from daily life problems (Saiz and Rivas 2008), may not be suitable to assess accurately the bachelor students' CT skills. Finally, three cycles in the activity might not be enough to trigger effective differences in the level of CT skills, or to perceive its differences. As referred in the literature, it is very important to nurture CT with repeated practice (Marin and Halpern 2011) favouring a crescendo of responsibilities, filing possible gaps, and mitigating the difficulties in understanding the objectives reported in few cycles' situations (Van Zundert, an, Sluijsmansd van Merrienboer 2010).

This research presented herein allowed us to clarify the potentialities and constraints of an online peer review methodology for CT development in future engineers. According to the students, the attractiveness of the tasks and ease of management of the activity along with the use of innovative tools and the collaborative writing in an anonymous way favoured their involvement in the tasks. From the teachers' perspective, this methodology promotes the organisation of activities enhancing the autonomous construction and communication of knowledge, as well as hetero and self-evaluation. Another important issue is that this type of activities does not involve costly platforms or advanced computer skills. Furthermore, it favours a student's pro-active attitude in lifelong learning.

In future work, we intend to deepen our investigation in several dimensions: (1) test the various scopes of the level X Cornell test (instead of only analysing its global performance), identify specific domains evidencing changes in the students' CT skills, or compare it to other available tests to establish the most suitable CT assessment tool for the students and the activity; (2) explore a longitudinal qualitative analysis of the students' assignments targeting the achieved competencies on individual base and (3) on a pedagogical perspective, deepen the teaching strategy and diversify the tools to be used in order to involve and motivate the students in the CT practice, including other Learning Management Systems platforms.

References

Angeli, C., and N. Valanides. 2009. "Instructional Effects on Critical Thinking: Performance on Ill-defined Issues." *Learning and Instruction* 19 (4): 322–334.

Arter, J., and J. Salmon. 1987. *Assessing Higher Order Thinking Skills. A Consumer's Guide*. Portland, OR: Northwest Regional Educational Laboratory.

Bauer, C., K. Figl, M. Derntl, P. P. Beran, and S. Kabicher. 2009. "The Student View on Online Peer Reviews." *ACM SIGCSE Bulletin* 41 (3): 26–30.
Behar-Horenstein, L. S. 2011. "Teaching Critical Thinking Skills in Higher Education: A Review of the Literature." *Journal of College Teaching & Learning* 8 (2): 25–42.
Bressy, G., and C. Konkuyt. 2008. *Management et économie des entreprises*. Chapitre 9, 9e ed. Phuket, Thailand: Sirey.
Broadbear, J. T. 2003. "Essential Elements of Lessons Designed to Promote Critical Thinking." *Journal of The Scholarship of Teaching and Learning* 3 (3): 1–8.
Brodahl, C., S. Hadjerrouit, and N. K. Hansen. 2011. "Collaborative Writing with Web 2.0 Technologies: Education Students' Perceptions." *Journal of Information Technology Education: Innovations in Practice* 10: 73–103. http://www.jite.org/documents/Vol10/JITEv10IIPp073-103Brodahl948.pdf
Brodie, P., and K. Irving. 2007. "Assessment in Work-based Learning: Investigating a Pedagogical Approach to Enhance Student Learning." *Assessment & Evaluation in Higher Education* 32 (1): 11–19.
Browne, M. N., and K. Freeman. 2000. "Distinguishing Features of Critical Thinking Classrooms." *Teaching in Higher Education* 5 (3): 301–309.
Butler, H. A., C. P. Dwyer, M. J. Hogan, A. Franco, S. F. Rivas, and L. S. Almeida. 2012. "The Halpern Critical Thinking Assessment and Real-world Outcomes: Cross-National Applications." *Thinking Skills and Creativity* 7 (2): 112–121.
Calvo, R. A., S. T. O'Rourke, J. Jones, K. Yacef, and P. Reimann. 2011. "Collaborative Writing Support Tools on the Cloud." *IEEE Transactions on Learning Technologies* 4 (1): 88–97.
Carpenter, C. B., and J. C. Doig. 1988. "Assessing Critical Thinking Across the Curriculum." *New Directions for Teaching and Learning* 1988 (34): 33–46.
Claris, L., and D. Riley. 2012. "Situation Critical: Critical Theory and Critical Thinking in Engineering Education." *Engineering Studies* 4 (2): 101–120.
Cooney, E., K. Alfrey, and S. Owens. 2008. "Critical Thinking in Engineering and Technology Education: a Review." 115th ASEE annual conference and exposition, Pittsburg, USA, June 22–25.
Cromwell, L. S. 1992. "Assessing Critical Thinking." In *Critical Thinking: Educational Imperative*, edited by C. Barnes, 37–50. San Francisco, CA: Jossey-Bass.
Cruz, G., C. Dominguez, A. Maia, D. Pedrosa, and G. Grams. 2013. "Web-based Peer Assessment: A Case Study with Civil Engineering Students." *International Journal of Engineering Pedagogy (IJEP), 3(S1):* 64–70.
Deal, K. H., and J. Pittman. 2009. "Examining Predictors of Social Work Students' Critical Thinking Skills." *Advances in Social Work* 10 (1): 87–102.
Dominguez, C., A. Maia, D. Pedrosa, and G. Grams. 2013. "Web-based Peer Assessment: A Case Study with Civil Engineering Students." *International Journal of Engineering Pedagogy* 3 (1): 64–70.
Ennis, R. 1993. "Critical Thinking Assessment." *Theory into Practice* 32 (3): 179–186.
Ennis, R. 1996. *Critical Thinking*. Upper Saddle River, NJ: Prentice Hall.
Ennis, R. 1997. "Inquiry: Critical Thinking across the Disciplines." *Spring* 16 (3): 1–9.
Ennis, R. 2003. "Critical Thinking Assessment." In *Critical Thinking and Reasoning*, edited by D. Fasko, 293–310. Cresskill, NJ: Hampton Press.
Ennis, R., and M. Goldman. 1991. "Critical Thinking: A Streamlined Conception." *Teaching Philosophy* 14 (1): 5–24.
Ennis, R., and J. Millman. 1985. *Cornell Critical Thinking Test, Level X*. Pacific Grove, CA: Midwest Publications.
Facione, P. A. 1990. *The California Critical Thinking Skills Test – College Leve. Technical Report #1: Experimental Validation and Content Validity*. Millbrae: California Academic Press.
Facione, P. A. 2011. *Critical Thinking: What It Is and Why It Counts*. San Jose: California Academic Press.
Fahim, M., and G. Nazari. 2012. "Practicing Action Research for enhancing Critical Thinking." *Journal of Science* 2 (1): 84–89.
Fani, T. 2011. "Overcoming Barriers to Teaching Critical Thinking." 1st international conference on the Future of Education, Florence, Italy, June 16–17.
Froyd, J. E., P. C. Wankat, and K. A. Smith. 2012. "Five Major Shifts in 100 Years of Engineering Education." *Proceedings of the IEEE* 100: 1344–1360.
Greenlaw, S. A., and D. B. Stephen. 2003. "Teaching Critical Thinking with Electronic Discussion." *The Journal of Economic Education* 34 (1): 36–52.
Halpern, D. F. 2010. *Halpern Critical Thinking Assessment*. SCHUHFRIED (Vienna Test System).
Hattie, J., and H. Timperley. 2007. "The Power of Feedback." *Review of Educational Research* 77 (1): 81–112.
Jamison, J. R. 2005. "Fostering Critical Thinking Skills: A Strategy for Enhancing Evidence Based Wellness Care." *Chiropractic & Osteopathy* 13 (19): 19–27.
Kang, N., and C. Howren. 2004. "Teaching for conceptual understanding." *Science and Children* 42 (1): 28–32.
Karandinou, A. 2012. "Peer-assessment as a Process for Enhancing Critical Thinking and Learning in Design Disciplines." *Transactions* 9 (1): 53–67.
Knight, L., and T. Steinbach. 2011. "Adapting Peer Review to an Online Course: An Exploratory Case Study." *Journal of Information Technology Education* 10: 81–100.
Krippendorf, K. 2013. *Content Analysis. An Introduction to its Methodology*. 3rd ed., Thousand Oaks, CA: Sage.
Ku, K. Y. 2009. "Assessing Students' Critical Thinking Performance: Urging for Measurements Using Multi-response Format." *Thinking Skills and Creativity* 4: 70–76.
Lai, E. R. 2011. *Critical Thinking: A Literature Review*. Pearson's Research Reports Accessed October 12, 2013. http://www.criticalthinking.org/aboutCT/define_critical_thinking.cfmhttp://images.pearsonassessments.com/images/tmrs/CriticalThinkingReviewFINAL.pdf

Linn, R. L., and N. E. Gronlund. 2000. *Measurement and Assessment in Teaching*. 8th ed. Englewood Cliffs, NJ: Merrill/Prentice Hall.
Litzinger, T., L. R. Lattuca, R. Hadgraft, and W. Newstetter. 2011. "Engineering Education and the Development of Expertise." *Journal of Engineering Education* 100 (1): 123–150.
Lu, R., and L. Bol. 2007. "A Comparison of Anonymous Versus Identifiable e-Peer Review on College Student Writing Performance and the Extent of Critical Feedback." *Journal of Interactive Online Learning* 6 (2): 100–115.
Marin, L. M., and D. F. Halpern. 2011. "Pedagogy for Developing Critical Thinking Adolescents: Explicit Instruction Produces Greatest Gains." *Thinking Skills and Creativity* 6 (1): 1–13.
Mwalongo, A. 2011. "Teachers Perceptions about ICT for Teaching, Professional Development, Administration and Personal Use." *International Journal of Education using Information and Communication Technology* 7 (3): 36–49.
Nair, C., A. Patil, and P. Mertova. 2009. "Re-engineering Graduate Skills – A Case Study." *European Journal of Engineering Education* 34 (2): 131–139.
Nelson, M. M., and C. D. Schunn. 2009. "The Nature of Feedback: How Different Types of Peer Feedback Affect Writing Performance." *Instructional Science* 27 (4): 375–401.
Norris, S. P. 1989. "Can We Test Validly for Critical Thinking?" *Educational Researcher* 18 (9): 21–26.
Norris, S. P. 1992. "Introduction: The Generalizability Question." In *The Generalizability of Critical Thinking*, edited by S. P. NorriSbibs, 1–15. New York: Teachers, College Press.
Norris, S. P., and R. H. Ennis. 1989. *Evaluating Critical Thinking*. Pacific Grove, CA: Critical Thinking & Press Software.
Oliveira, M. 1992. "A criatividade, o pensamento crítico e o aproveitamento escolar em alunos de ciências." PhD Thesis, University of Lisbon: Faculty of Sciences. Lisbon, Portugal.
Ozogul, G., and H. Sullivan. 2009. "Student Performance and Attitudes under Formative Evaluation by Teacher, Self and Peer Evaluators." *Educational Technology Research and Development* 57 (3): 393–410.
Paul, R., and L. Elder. 2003. *The Miniature Guide to Critical Thinking: Concepts and Tools*. 3rd ed. Dillon Beach, CA: The Foundation for Critical Thinking Press.
Pithers, R. T., and R. Soden. 2000. "Critical Thinking in Education: A Review." *Educational Research* 42 (3): 237–249.
Ralston, P., and C. Bays. 2010. "Refining a Critical Thinking Rubric for Engineering." 117th ASEE annual conference and exposition, Kentucky, USA, June 20–23.
Rienzo, T., and B. Han. 2009. "Microsoft or Google Web 2.0 Tools for Course Management." *Journal of Inform ation Systems Education* 20 (2): 123–127.
Saiz, C., and S. Rivas. 2008. "Assessment in Critical Thinking: A Proposal for Differentiating Ways of Thinking." *Ergo Nueva Epoca* 22–23: 25–26.
Scriven, M., and R. Paul. 2007. *Defining Critical Thinking*. The Critical Thinking Community: Foundation for Critical Thinking. Accessed October 12, 2013. http://www.criticalthinking.org/aboutCT/define_critical_thinking.cfm
Sgro, S., and S. A. Freeman. 2008. "Teaching Critical Thinking Using Understanding by Design." Proceedings of the 2008 ASEE annual conference & exposition, Pittsburgh, PA, June 22–25.
Shuman, L. J., M. Besterfield-Sacre, and J. McGourty. 2005. "The ABET "Professional Skills" – Can They Be Taught? Can They Be Assessed?" *Journal of Engineering Education* 94 (1): 41–55.
da Silva Nascimento, M. 2014. "Come Together: Peer Review with Energy Engineering Students." *International Journal of Engineering Pedagogy* 4 (5): 34–41.
Sondergaard, H. 2009. "Learning from and with Peers: The Different Roles of Student Peer Reviewing." *ACM SIGCSE Bulletin* 41 (3): 31–35.
Stein, B., A. Haynes, M. Redding, T. Ennis, and M. Cecil. 2007. "Assessing Critical Thinking in STEM and Beyond." In *Innovations in E-learning, Instruction, Technology, Assessment, and Engineering Education*, edited by M. Iskander, 79–82. Dordrecht, The Netherlands: Springer.
Strijbos, J.-W., S. Narciss, and K. Dünnebier. 2010. "Peer Feedback Content and Sender's Competence Level in Academic Writing Revision Tasks: Are They Critical for Feedback Perceptions and Efficiency?" *Learning and Instruction* 20: 291–303.
Sung, Y.-T., K.-E. Chang, S.-K. Chiou, and H.-T. Hou. 2005. "The Design and Application of a Web-based Self- and Peer-Assessment System." *Computers & Education* 45 (2): 187–202.
Tenreiro-Vieira, C. 2004. "Produção e avaliação de actividades de aprendizagem deciências para promover o pensamento crítico nos alunos." *Revista Iberoamericana de Educación* 33: 1–17.
Tsui, L. 2002. "Fostering Critical Thinking through Effective Pedagogy: Evidence from Four Institutional Case Studies." *Journal of Higher Education* 73 (6): 740–763.
Van Zundert, M., D. Sluijsmans, and J. van Merriënboer. 2010. "Effective Peer Assessment Processes: Research Findings and Future Directions." *Learning and Instruction* 20: 270–279.
Xiao, Y., and R. Lucking. 2008. "The Impact of Two Types of Peer Assessment on Students' Performance and Satisfaction within a Wiki Environment." *The Internet and Higher Education* 11 (3–4): 186–193.
Walker, S. E. 2003. "Active Learning Strategies to Promote Critical Thinking." *Journal of Athletic Training* 38 (3): 263–267.

About the authors

Caroline Dominguez is a researcher at CIDTFF and has been a lecturer at the University of Trás-Montes e Alto Douro (UTAD) since 1995, in Vila Real, Portugal, where she teaches business management. Her main research interests are related to educational issues, project and quality management, and human resources. Before pursuing an academic career,

she was the head manager of organisations dealing with development issues in Peru and Portugal, where she developed and implemented various international funded projects.

Maria M. Nascimento is a researcher at Lab DCT-UTAD/CIDTFF and has been a lecturer at the University of Trás-Montes e Alto Douro (UTAD) since 1985, in Vila Real, Portugal, where she teaches Statistics, Sampling, Experimental Design and Operations Research. Her main research interests are related to teaching Statistics and its Attitudinal and Didactical issues, as well as the Ethnomathematics research field. Recently the critical thinking research field was discovered in its connections to statistical thinking.

Rita Payan-Carreira is a researcher at CECAV and has been an assistant Professor at the University of Trás-Montes e Alto Douro (UTAD), in Vila Real, Portugal, in the field of Veterinary Medicine (animal reproduction and pathology). Her research interests also cover educational issues, project and quality management, and human resources.

Gonçalo Cruz is a research fellow at INESC TEC (formerly INESC Porto) and the University of Trás-os-Montes e Alto Douro (UTAD), in Portugal, where he works as e-learning consultant since 2010. In 2011, Gonçalo concluded his Masters in Educational Sciences by the University of Coimbra. His main research interests are Distance Learning, MOOCs, Social Learning, Technology Enhanced Learning, CSCL, E-assessment, Educational Psychology, Virtual Worlds and e-learning maturity models. Currently, Gonçalo is a Ph.D. student in Didactics of Science and Technology at UTAD.

Helena Silva is an associate professor at the University of Trás-Montes e Alto Douro (UTAD), in Vila Real, Portugal, since 2001. Her main research and teaching interests are related to teaching methodologies and teachers' professional development, with emphasis on cooperative learning, formative assessment and communities of practice. She works with educators who seek to create classrooms that are more effective in academics and social skills. Recently the critical thinking research field was discovered in its connections to professional development. She is also an educational author.

José Lopes is an associate professor at the University of Trás-Montes e Alto Douro (UTAD), in Vila Real, Portugal, since 1998. His main research and teaching interests are related to teaching educational psychology and teachers' professional development, with emphasis on cooperative learning, formative assessment, and communities of practice. He works with educators who seek to create classrooms that are more effective in academics and social skills. Recently the critical thinking research field was discovered in its connections to professional development. He is also an educational author.

Maria da Felicidade A. Morais is a researcher at CEL-UTAD and has been a lecturer at the Universidade de Trás-os-Montes e Alto Douro (UTAD), since 1993. Her main research interests are in the area of Textual Linguistics, Pragmatics, and Teaching of Languages. In working with future media professionals, she has been promoting the use of peer review tasks, aiming at the improvement of critical thinking and communication proficiency.

Eva Morais is a researcher at CEMAPRE and CM-UTAD, and has been a lecturer at the University of Trás-os-Montes e Alto Douro (UTAD), in Vila Real, Portugal, since 2000, where she teaches Probability and Statistics courses. In 2013, she concluded her Ph.D. in Mathematics Applied to Economy and Management by ISEG, at the Technical University of Lisbon. Her main interests are related to the study of methods used to solve partial differential equations appearing in financial pricing models. Recently, Eva Morais has become interested in the critical thinking research field.