THINK4JOBS GUIDELINES

A protocol for Critical Thinking transfer from curricula to labour market

THINK4JOBS

Guidelines:

A protocol for Critical Thinking transfer from curricula to labour market

Co-funded by the Erasmus+ Programme of the European Union

Technical and Cataloguing Data

Cover layout: Designed in Canva

Date of Publication: 2023

Recommended Citation: Payan Carreira, R., Rebelo, H., Sebastião, L., Sacau, A., Ferreira, D., Simões, M., Pnevmatikos, D., Christodoulou, P., Lithoxoidou, A., Georgiadou, T., Papadopoulou, P., Spyrtou, A., Papanikolaou, A., Oikonomou, A., Dumitru, D., Mihăilă, R., Badea, L., Minciu, M., Kriaučiūnienė, R., (…) Paun, D. (2023). *THINK4JOBS Guidelines: A protocol for Critical Thinking transfer from curricula to labour market.* Greece: University of Western Macedonia. ISBN: 978-618-5613-11-2. URL: *https://think4jobs.uowm.gr/results/intellectualoutput4*

ISBN: 978-618-5613-11-2

Funding: This work has been supported by the "Critical Thinking for Successful Jobs - Think4Jobs" Project, with the reference number 2020-1-EL01-KA203-078797, funded by the European Commission/EACEA, through the ERASMUS+ Programme.

Co-funded by the Erasmus+ Programme of the European Union

Authors

1. Payan Carreira Rita, University of Évora (UÉvora)
2. Rebelo Hugo, University of Évora (UÉvora)
3. Sebastião Luís, University of Évora (UÉvora)
4. Sacau Ana, consultant at the University of Évora (UÉvora)
5. Ferreira David, University of Évora (UÉvora)
6. Simões Margarida, University of Évora (UÉvora)
7. Pnevmatikos Dimitrios, University of Western Macedonia (UOWM)
8. Christodoulou Panagiota, University of Western Macedonia (UOWM)
9. Lithoxoidou Angeliki, University of Western Macedonia (UOWM)
10. Georgiadou Triantafyllia, University of Western Macedonia (UOWM)
11. Penelope Papadopoulou, University of Western Macedonia (UOWM)
12. Spyrtou Anna, University of Western Macedonia (UOWM)
13. Papanikolaou Anastasios, University of Western Macedonia (UOWM)
14. Oikonomou Anastasia, University of Western Macedonia (UOWM)
15. Dumitru Daniela, Bucharest University of Economics Studies (ASE)
16. Mihăilă Robert, Bucharest University of Economics Studies (ASE)
17. Badea Liana, Bucharest University of Economics Studies (ASE)
18. Minciu Mihaela, Bucharest University of Economics Studies (ASE)
19. Kriaučiūnienė Roma, Vilnius University (VU)
20. Ivancu Ovidiu, Vilnius University (VU)
21. Poštič Svetozar, Vilnius University (VU)
22. Arcimavičienė Liudmila, Vilnius University (VU)
23. Vaidakavičiūtė Agnė, Vilnius University (VU)
24. Mäkiö Juho, University of Applied Sciences Emden-Leer (HSEL)
25. Mäkiö Elena, University of Applied Sciences Emden-Leer (HSEL)
26. Silva Ruben, Hospital Veterinário Atlântico (HVA)
27. Miranda Sonia, Hospital Veterinário Atlântico (HVA)
28. Kappatou Anastasia, Elementary Experimental School of Florina
29. Sechidis Kostantinos, Elementary Experimental School of Florina

Co-funded by the Erasmus+ Programme of the European Union

30. Amarantidou Kiriaki, Elementary Experimental School of Florina
31. Arvanitakis Ioannis, Elementary Experimental School of Florina
32. Doukas Dimitrios, Elementary Experimental School of Florina
33. Antonogianni Vasiliki, Elementary Experimental School of Florina
34. Auškelienė Audronė, Public Service Language Center (VIKC)
35. Rudienė Asterija, Public Service Language Center (VIKC)
36. Samukienė Rita, Public Service Language Center (VIKC)
37. Busker Wolfgang, Orgadata AG (Orgadata)
38. Meinders Andreas, Orgadata AG (Orgadata)
39. Maioru Monica, BRD Groupe Société Générale (BRD)
40. Paun Diana, BRD Groupe Société Générale (BRD)

Co-funded by the Erasmus+ Programme of the European Union

Table of Contents

Co-funded by the Erasmus+ Programme of the European Union

Co-funded by the Erasmus+ Programme of the European Union

Executive summary and key findings

The Intellectual Output 4 (IO4) reports the results from the implementation of the Critical Thinking (CT) Blended Apprenticeships Curricula (CTBAC) described in the third Intellectual Output IO3 [1], and it discusses the recorded gains in CT skills and dispositions in students enrolled in the piloting activities. A cross- and intra-disciplinary analysis, resulting from the comparison between the scores obtained before and at the end of the piloting CTBACs courses, provides support to the recommendations proposed by the partnership for the Critical Thinking Blended Apprenticeships Curricula implementation, which are gathered under the part III of this report: "THINK4JOBS guidelines for CT transfer from curricula to apprenticeships".

University of Évora (UÉvora), Portugal was the partner who led the delivery of the IO4. The objectives of IO4 were defined as follows:

1. Assess the changes in CT skills and dispositions associated with the implementation of 12 CT blended apprenticeship curricula as developed by the University-Business partnership for the disciplines Business Informatics, Teacher Education, Veterinary Medicine, and Business and Economics and the course English as a Foreign Language;
2. Use this data for a cross-disciplinary analysis;
3. Present the "THINK4JOBS guidelines for CT transfer from curricula to apprenticeships".

The CTBACs implementation was at the inception of these objectives. The implementation of these curricula was facilitated by the close collaboration between Higher Education (HE) instructors and Labour Market Organisation (LMO) tutors in creating the scenarios. The Moodle platform was used as a learning interface for CTBACs. The implementation of the new curricula occurred at the fall and spring terms of the 2021/2022 academic year. Albeit CTABCs will be repeated in the 2022/23 academic year, the analysis presented in here does not consider this.

Co-funded by the Erasmus+ Programme of the European Union

In total, 609 students were enrolled in the piloting activities, a larger number than the initially envisaged in the project submission (150 students) (Table 1). Still, not all the participating students responded to the questionnaires. Respondents that filled the pre- and the last post-test questionnaires represented 54% of the students engaging in the activities. A difficulty arose to reach the initially proposed numbers of control students (non-engaged in the pilot courses) since most courses were not offered in two different semesters and students did not accept to be left out of activities that they perceive of bringing some benefits for their success. Only the Greece and Portuguese partners succeeded in organizing a control group, even though the one for Portugal was of a small size.

Table 1 – Pilot CT-blended courses implemented in the academic year of 2021/2022.

Country	Programme/Discipline	Courses	No. students registered in CTBACs	No. students in control courses
Germany	Business Informatics	Design patterns	14	- -
		Innovation Management	10	- -
		Economic Aspects of Industrial Digitalization	10	- -
		Scientific seminar	10	- -
Greece	Teacher Education	Teaching Biological concepts	83	- -
		Teaching Science Education	61	84
		Teaching of the Study of the Environment	12	- -
Lithuania	International Relations and Political Science	English as Foreign Language	61	
Portugal	Veterinary Medicine Integrated Masters	Imaging	78	- -
		Deontology	56	- -
		Gynecology, Andrology and Obstetrics	71	- -
		Curricular Traineeship	- -	12
Romania	Business and Economics	Business communication	69	- -
		Pedagogy and Didactics of Financial Accounting	48	- -
		Virtual Learning Environments in Economics	26	- -
Total			**609**	**96**

The gains in students' CT skills and dispositions following the CTBACs implementation were assessed using a new instrument that merged two questionnaires [a short-form of the Nair's Critical Thinking Self-assessment Scale (CTSAS) [2] developed for this specific purpose, and the Student-Educator Negotiated Critical Thinking Dispositions Scale (SENCTDS) [3], which was applied to the students before, during and after the interventions. The version of the instrument in the original language (English) was translated into German, Greek, Romanian

Co-funded by the Erasmus+ Programme of the European Union

and Portuguese, to be implemented by the partners. The two questionnaires were merged into one google form document, for commodity of the respondents and to ensure that both were filled at the same time.

The validation of the instruments was performed using the answers gathered from all the linguistic versions of the questionnaires. The preliminary independent validation of both questionnaires showed that they represent a strong tool, with a good goodness-of-fit indices and a strong internal consistency. The invariance analysis confirmed that both the part of the instrument that assessed CT skills and the CT dispositions remained stable between countries, therefore supporting the quality of the instrument and of the translated versions used (German, Greek, Romanian and Portuguese vs. English). These versions represent additional added-value productions from the project.

The cross-disciplinary analysis evidences the existing gains associated with the implementation of the blended curricula. Some baseline differences were found by countries, age and sex in the overall population for specific CT skills and dispositions. These changes may reflect the background cultural or experiential differences in students from the various disciplines/courses involved in the project [Germany (Business Informatics), Greece [Teacher Education], Lithuania (course of English as a Foreign Language), Romania (Business and Economics), and Portugal (Veterinary Medicine)].

CTBACs-related gains were more evident in skills than dispositions (ca. ten points vs. one point in the respective integrated scores), which may derive from the fact that it is more difficult to change attitudes (dispositions) in short-timed interventions than procedures. Moreover, even though the control groups were not possible for all the countries, when they existed results show higher gains in the integrated scores of CT skills and dispositions for students enrolled in CTBACs than in students engaged in the control group, particularly in the *Evaluation*, *Inference* and *Explanation skills*, and in the disposition *Attentiveness*.

From the interpretation of the results gathered in the implementation of the blended apprenticeship curricula, the guidelines for implementation of CTBACs were prepared around the following steps:

Co-funded by the Erasmus+ Programme of the European Union

1. Explain what you are doing – Explain to the students why critical thinking is a crucial competency in today workforce, and how it is understood and praised in a particular profession. Explicitly include the development of CT within the outcomes of your course.
2. CT training must be a continuous and pervasive process. – To succeed, an effort is needed across the discipline curricula to endorse students' CT skills and dispositions; skills need training and dispositions require internalization of the desirable attitudes, so time and a combined effort at the disciplinary level are necessary to obtain more and consistent gains or positive changes.
3. Get time to do it – supporting the previous item, time is needed also at the course level to work on the proposed goals. Both students and Educators need to schedule the activities, so they have time to prepare, develop, and provide or receive feedback regarding the students´ performance, so the intervention leads to meaningful learning.
4. Get connected to reality – by presenting students with cases issued from situations professionals face daily, students' motivation increases, and they perceive more positively the learning experiences.
5. Accept reasonable risk – development of critical thinking may benefit from presenting students with complex problems with uncertain solutions, where students are allowed to stumble, since failure (an incorrect decision-making) under a safe environment, allows students to think on the premises that drove to the error and corrective feedback from the educator plays a crucial role as a learning tool.
6. Reflect on CT skills and dispositions development – offer the students the opportunity to reflect on the changes on the way they reasoned through the situations or the attitudes they developed in order to increase the effect of the learning interventions and better cultivate reflective thinking about one' experiences.

Co-funded by the Erasmus+ Programme of the European Union

Introduction

According to OCDE, an increasing gap exists between the qualifications certified by Higher Education Institutions (HEI), and the generic, 21st-century skills requested by Labour Market Organizations (LMO), particularly on what concerns literacy and critical thinking skills encompassing problem solving, analytic reasoning and communication [4]. According to employers, CT fosters the conditions for a person to continuously improve their reasoning in order to adapt quickly to organizational change [5]. At the same time, employers believe that critical thinking enables employees to adapt more quickly to the challenges of a changing environment and to find the best solutions for the organisation, customers and themselves [6]. Pondering the role of the HEI in formal professional education and the volume of graduates arriving each year to the labour market, this statement can be disturbing, since it can be understood that, today, a tertiary qualification may not fully guarantee graduates higher skills for a particular profession. In part, this issue may emerge from a context where the skills' demand is rapidly changing to match the constant high pace driven by the fast growth in technological and scientific knowledge.

This awareness has driven higher education to implement reforms, which shifted the learning approach from a lecture format to a student-centred one, requesting the active participation of students in the learning process. In that way, students are required to transpose the acquired knowledge into the solution of new or different problematic situations [7], to support their decision-making process and search for new forms to solve a situation or address a problem.

Asymmetries in the newly graduates' competencies has been at the focus of Think4Jobs Project, tackling the issue in a joint collaborative approach between HEI and LMO in the design of Critical Thinking Blended Apprenticeships Curricula (CTBAC) for the development of Critical Thinking (CT) skills and dispositions implementing blended interventions within some courses of a particular discipline [8, 9].

The CTBACs were implemented using the e-learning platform Moodle. This platform allowed teachers and stakeholders to deploy the interventions in each course separately and support

Co-funded by the Erasmus+ Programme of the European Union

the participation of learners enrolled in the courses. The platform further acted as a repository of the material provided during the courses and registered the results (grades) of the students in the learning activities. The implementation of the learning activities followed the design proposed in IO3 [1]. An important question the Project proposed to investigate is whether the new instructional design has contributed to students´ CT improvement at the end of the pilot courses, as it was anticipated in the course design, and to ascertain the relative progress or the short-term "learning gain' in terms of CT skills and dispositions. To collect the necessary information, a pre-test/post-test approach was selected, using as means for data collection an instrument that merged down two questionnaires, one addressing the CT skills and the other CT dispositions.

This Intellectual Output – IO4 – aims at:

1. Identifying changes in the CT skills and dispositions in students enrolled in the CT blended apprenticeship curricula, based on the comparison of scores obtained in each Course/Discipline at a pre-test and post-test moments;
2. To evaluate the results of the implementation of the new learning interventions and discuss them to identify the need for putative modifications of the learning scenarios;
3. To present the "THINK4JOBS guidelines for CT transfer between HEI and LMOs.

The presentation of the information in IO4 is divided in three main sections, starting with the selection and validation of the questionnaires used for scoring CT skills and dispositions, followed by the presentation of cross and intra-disciplinary data analysis and the interpretation of the results, which support the recommendations presented as the THINK4JOBS guidelines for CTBACs implementation.

Co-funded by the Erasmus+ Programme of the European Union

PART I – VALIDATION OF THE INSTRUMENTS USED TO ASSESS THE STUDENTS' CT SKILLS AND DISPOSITIONS

1. Why and how to evaluate CT changes in students after piloting CTBAC

Evaluation of educational interventions is crucial to assess the success of changes introduced in the learning-process, or in the curricula. It must be integrated in an evaluation cycle planned to demonstrate that the intervention reached the intended objectives [10]. According to Wilkes and Bligh [10], the evaluation should cover the acquisition of skills, knowledge, and attitudes, directed to the proposed learning outcomes, so as to drive and support curricular change.

In the Think4Jobs Project, it was decided to use a student-oriented approach aiming at the enhancement of CT skills and dispositions within a course specific context, and the new curricula for the courses was presented in an earlier Project output [1]. The effectiveness of the learning interventions in fostering CT competences was evaluated using two questionnaires (the CTSAS-SF and SENCTDS), aligned with the CT-related learning goals proposed for each course (for details, see [1]), in a short-term assessment approach.

In the literature, there are multiple formal standardized CT tests (such as the CCTT - Cornell Critical Thinking Test; the California Critical Thinking Dispositions Inventory - CCTDI; or the Halpern Critical Thinking Assessment test- HCTA, among others) [11] that could be used to assess CT, even if these instruments more frequently address the assessment of skills than dispositions. Still, the construct transferability of many standardized CT tests to different populations or disciplines has been questioned [12, 13]. Together with the multiple conceptualizations of CT, this might explain why an agreement about a standard instrument used to evaluate consistently learning effectiveness remains to be reached. Besides, new instruments keep being developed and assayed across disciplines and cultural contexts [10, 11].

Additional drawbacks have been associated with the standardized CT tests: they are not easily accessed everywhere; some require expert evaluation and scoring and the training of ratters

Co-funded by the Erasmus+ Programme of the European Union

to minimize individual bias [14]. Some of them are too expensive to be routinely applied [15], while others present situations that are inauthentic and distant from the students' reality [11, 12]. Also, standard tests tend to be focused particularly on skills or address skills and dispositions combined, and they are usually long, requiring between 50 to 80 minutes to be completed [12], driving to poor motivation of students to fill the questionnaires [16]. Moreover, for some tests, it is unclear how the respondent's reasoning will identify more discrete dispositions, such as open-mindedness or inquisitiveness [11], or how they distinguish it from the application of disciplinary specific reasoning skills gathered across the programme by students enrolled in more advanced years of their academic path.

Self-reported questionnaires to assess students' perceptions of CT have been developed and validated in the past decades for different disciplines, even if their use remains non-consensual [17, 18], particularly due to the poor reliability of the constructs if used in different populations or disciplines. Despite the controversy, self-report questionnaires often are used to assess CT perceived changes after the implementation of new instructional methods. In this context, they may present advantages in assessing non-cognitive competencies, particularly when they don´t have a direct reflection on the students' grades, serving just with the purpose of monitoring or enhancing students' performance and to identify individual training needs [19].

Regarding the assessment of CT dispositions, available instruments are scarce [20, 21] and the few available are grounded on different conceptualizations of the attitudinal dimensions, according to the context of the studies. As it happens with instruments measuring CT skills, some of the scales lack consistency [20].

Consequently, the Think4Jobs consortium decided to apply a self-report questionnaire for the facility of its use, the need to use it repeatedly across a course, and its friendly use by students across disciplines and countries. Since the Facione framework was used to identify the CT skills and dispositions targeted in the educational interventions, as exposed in IO3 [1], the consortium restricted the search to questionnaires targeting preferably the assessment of CT skills and dispositions as conceptualized by Facione framework [22, 23].

Co-funded by the Erasmus+ Programme of the European Union

2. Methodology

Considering that the project intends to measure skills and dispositions associated with CT, and because skills and dispositions are seldom evaluated separately but with one instrument, the project consortium decided to resource two different instruments (one covering CT skills and the other the dispositions), to be merged within a unique instrument.

Based on the partners previous experience and the available literature, some primary criteria were established to select the instrument: to be a closed-end test; easy to administer online; matching the proposed CT skills and dispositions identified as outcomes for the activities in the pilot courses; practical for students in taking; and not demanding in terms of the level of technical expertise required to answer and to retrieve information from. Moreover, the merged questionnaire should meet an additional criterion, namely the time to completion should be limited to 60 minutes, to avoid desponding students' willingness in using the tool.

CTSAS-SF, the tool used to assess CT skills in the Think4Jobs Project

Among the available instruments to assess CT skills under the conceptualization of Facione, the consortium selected the Critical Thinking Self-Assessment Scale (CTSAS) developed by Nair [24] for HE students. The questionnaire has been tested in different geographic and cultural contexts (which was considered a strength of the instrument) and scored well in the reliability and internal consistency tests, as well as in the confirmatory factor analysis [25]. The original CTSAS questionnaire was composed of 115 items covering six dimensions scored according to a seven-point rating scale (ranging from 0=never to 6=always) and was considered too long to complete (the author reported about 50 min), particularly because it was planned to piece it together with another instrument designed to score CT dispositions. Therefore, it was decided to use a short form of the CTSAS questionnaire, developed specially for the Project.

The original Nair' scale was shortened using a two-step approach under the expertise of two Portuguese researchers, by applying the following criteria for item rejection: 1. The items with loading weights below .500 were eliminated, remaining 84 items; 2. Redundant items and those with a non-cognitive focus were marked for elimination, leaving 58 items. After marking

Co-funded by the Erasmus+ Programme of the European Union

the items to eliminate, the remaining construct was analysed by two independent experts to confirm or revert the rejecting proposal based on the Facione' conceptualization of CT skills and subskills. These experts accepted most deletions, but recommended the retention of items 16 and 19 from the original scale, due to their theoretical relevance. At the end of the process, the CTSAS short-form retained in total 60 peer-reviewed items. No modifications were introduced to the items retained of the original CTSAS scale (for details see [2]).

In the CTSAS short form (CTSAS-SF), the number of items assessing each dimension ranged between 7 and 13. For subdimensions (or subskills), the number of items varied between 3 to 7 items, with exception for 5 subdimensions (*decoding significance*, *detecting arguments*, *assessing claims*, *stating results*, and *justifying procedures*), which included only two items each (Table 2). The CTSAS-SF maintained the original scale's framework, where students start from the question "*What do you do when presented with a problem?*" and are requested to answer the items using a seven-point Likert scale structure with the following correspondence: 0= Never; 1 = Rarely; 2 = Occasionally; 3 = Usually; 4 = Often; 5 = Frequently; 6 = Always.

Table 2 – The structure and dimensions of the CTSAS-SF questionnaire

	CTSAS Dimensions (Skills / Subskills)	Items in the CTSAS short-form
Interpretation	Categorization	1 – 3
	Clarifying meaning	6 – 9
	Decoding significance	4, 5
Analysis	Detecting arguments	15, 16
	Analyzing arguments	17 – 20
Evaluation	Examining ideas	10 – 14
	Assessing claims	21, 22
	Assessing arguments	23 – 27
Inference	Drawing conclusions	36 – 40
	Conjecturing alternatives	31 – 35
	Querying evidence	28 – 30
Explanation	Stating results	41, 42
	Justifying procedures	43, 44
	Presenting arguments	45 – 50
Self-regulation	Self-examination	51 – 57
	Self-correction	58 – 60

Co-funded by the Erasmus+ Programme of the European Union

SENCTDS, the tool used to assess CT dispositions in the Think4Jobs Project

As already mentioned, instruments to measure CT dispositions are far less abundant than those assessing CT skills. Among the available questionnaires addressing the evaluation of CT dispositions, the consortium adopted the Student-Educator Negotiated Critical Thinking Dispositions Scale (SENCTDS), developed by Quinn et al. [3], which was validated in a mixed Irish and American student population. The scale was designed considering a different set of CT dispositions that the authors considered as being important for the labor market and to real world decision-making outcomes [3]. Some items in the scale combine some of the classical Facione CT dispositions into new dimensions foreseen as important for the academic and labor market success (e.g., in the dispositions *Organization*, *Perseverance* and *Intrinsic Goal Motivation*). The items represent six dispositions' dimensions (*Reflection*, *Attentiveness*, *Open-mindedness*, *Organization*, *Perseverance*, and *Intrinsic Goal Motivation*), and are phrased as statements against which the students must position themselves using a 7-point Likert scale: 1 = Strongly Disagree; 2 = Disagree; 3 = Slightly Disagree; 4 = Neither Agree nor Disagree; 5 = Slightly Agree; 6 = Agree; 7 = Strongly Agree. A strong point in favor of the questionnaire is that it correlates well with other validated constructs intent to evaluate CT dispositions [3]. The SENCTDS was used in its original version, which comprises of 21 items (Table 3). Despite the slight difference on the identification of the scale points (ranging from 1 = Strongly Disagree to 7 = Strongly Agree) compared to that of the CTSAS-SF (ranging from 0=never to 6=always), it was decided to retain the original representation of the Likert scale.

Translation of the CT skills and disposition instruments

The adopted CTSAS_SF and SENCTDS, originally in English, were translated into Portuguese, Romanian, Greek and German languages. The translation into these languages followed the recommended procedures (translation, revision, and refinement) to ensure that the meaning, connotation and conceptualization respected the original instrument [26, 27]. Two bilingual translators from each Country using a non-English version questionnaire, converted the adopted instruments into their mother language; different sets of researchers then analyzed the translations to screen differences between the two versions of the questionnaire and

Co-funded by the Erasmus+ Programme of the European Union

ensure the precision of the translation and its compliance with the original [28]. The translated versions were then reviewed by a group of experts from each national team in the project, who judged the content equivalence of the instrument. The experts' concordance was considered as an equivalent assessment of the translated questionnaire.

Table 3 – The structure and dimensions of the SENCTDS questionnaire

SENCTDS Dimensions (Dispositions)	Items in SENCTDS
Reflection	1 – 3
Attentiveness	4 – 7 (scored in reverse)
Open-mindedness	8 – 11 (scored in reverse)
Organization	12 – 14
Perseverance	15 – 17
Intrinsic Goal Motivation	18 – 21

Data collection and analysis

For the validation of the questionnaires all the first participants' responses in the blended courses piloted either during the fall/winter and the spring/summer terms were used in the five Higher Education Institutions (HEIs) participating in the project.

Responses from 531 university students (389 female, 142 male), with ages ranging from 19 to 58 years old (Mean = 23.47; SD = 7.184) were considered in this analysis. Figure 1 shows the age distribution in the population of respondents, and Figure 2 presents the unbalanced sex distribution per country.

Co-funded by the Erasmus+ Programme of the European Union

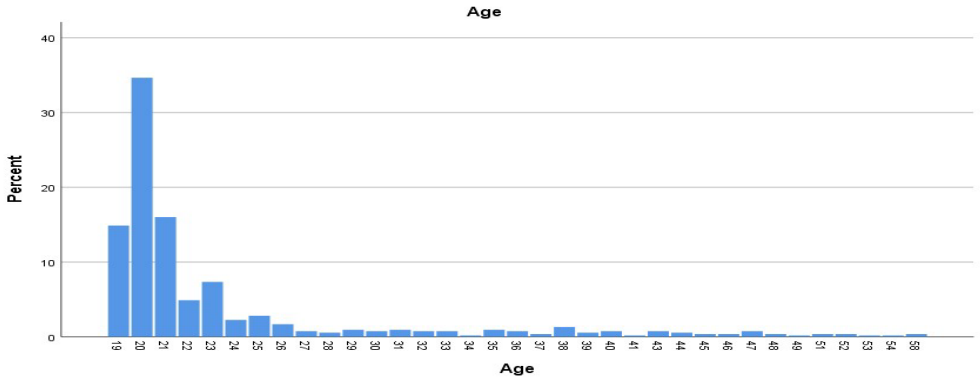

Figure 1 – Age distribution in the surveyed population (n=531)

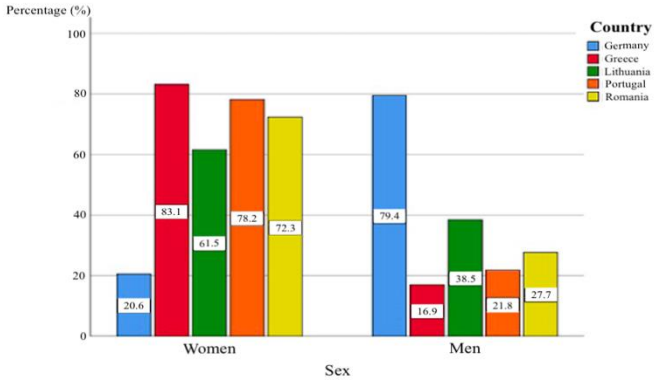

Figure 2 – Sex distribution *per* country (n=531).

An uneven distribution of participants by country was observed: 33.3% were from Greece, 29.4% from Portugal, 21.1% from Romania, 9.8% from Lithuania and 6.4% from Germany. Students responded to questionnaires in Greek, Portuguese, Romanian, English, and German, respectively. Students attended the following programmes (disciplines): Business Informatics (Germany), Teacher Education (Greece), Veterinary Medicine (Portugal) and Business and Economics (Romania), as well as the course English as Foreign Language (Lithuania).

Students signed an informed consent associated with the questionnaire and were informed that they could withdraw from the study at any time without penalty or loss of benefits. The study population represented a non-randomized, non-probability convenience sample resulting from the voluntary responses from students enrolled in the Think4Jobs' designed CTBACs.

Co-funded by the Erasmus+ Programme of the European Union

The CTSAS-SF and SENCTDS instruments were merged into one unique form that was made available on the Google Forms platform, upon an invitation sent via the Moodle page of the course students were enrolled in. The process was supervised by the teachers involved in the pilot courses. The form also contained a preliminary section regarding the general identification of students (email, name, country, discipline and course, sex, and age). The responses were recovered from the Google platform in an Excel file, *per* country. The email and name – necessary for pairing consecutive responses through time – were removed for the analysis of the results, during the anonymization step of the database preparation. The names were switched to an alpha-numeric code (composed by the code for the country - GR, LT, RO, GE, and PT respectively for Greece, Lithuania, Romania; Germany and Portugal - plus a sequential number, from 1 to *n*), and the column of the names and emails were deleted. A different researcher than the one plotting the statistical analysis conducted the data anonymization to reduce the risk of bias. Each country database was screened for inconsistent data, before merging them into a unique database for statistical analysis.

The statistical analysis included the Items descriptive measures (mean, standard deviation, skewness, kurtosis), the equal distribution Kolmogorov-Smirnov test and Mann-Witney's U for means' ranking differences. To assess if the CTSAS-SF and SENCTDS fitted the original factor model, a confirmatory factor analysis (CFA) was performed independently for each questionnaire, with weighted least square means and variances (WLSMV) as an estimation method due to the ordinal nature of data [29]. Model fit indices performed included the χ^2 test for exact fit, the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI) and the Root Mean Squared Error of Approximation (RMSEA). Following Hu and Bentler [30], we considered CFI and TLI values ≥ .90 and RMSEA ≤ .06 (90%IC) as acceptable fit values. Data were specified as ordinal in the model.

The reliability and internal consistency of the scale and subscales were estimated from Cronbach's Alpha. According to Hair et al [31], alphas above .70 were considered as good reliability indices. The multigroup invariance was assessed for sex and countries/disciplines. Differences between the RMSEA and CFI values under .015 and .01, respectively, were used as criteria for invariance [32, 33]. Univariate descriptive and internal consistency was

Co-funded by the Erasmus+ Programme of the European Union

calculated using IBM SPSS Statistics 26. CFA and multigroup invariance analysis were performed using MPlus 7.4 [34].

3. Results – validation of CT skills and Dispositions scales

The results will be presented divided into three parts, for both the CT skills and CT dispositions scales (CTSAS-SF and SENCTDS, respectively). The first section presents descriptive statistics of the items. The second section shows the results from the confirmatory factor analysis. The third section shows multigroup invariance analysis.

Descriptive analysis of items in CTSAS-SF and SENCTDS scales

For the CTSAS-SF, the mean range of the 60 items varies from 3.13 ("*I write essays with adequate arguments supported with reasons for a given policy or situation*") to 5.04 ("*I try to figure out the content of the problem*"). The standard deviation varies from .958 ("*I try to figure out the content of the problem*") to 1.734 ("*I write essays with adequate arguments supported with reasons for a given policy or situation*"). K-S test shows that data are equally distributed by female and male students ($p > .050$) except for item "*I can logically present results to address a given problem*" ($Z = 1.533$; $p = .018$) and item "*I respond to reasonable criticisms one might raise against one's viewpoints*" ($Z = 1.772$; $p = .004$). The item' description is displayed in the supplementary Table 2.

Mann-Witney's U test shows no statistically significant differences in CT skills (CTSAS-SF) between female and male students ($p > .050$) except for items "*I observe the facial expression people use in a given situation*" (Std U = -2.230; $p = .026$), "*I can logically present results to address a given problem*" (Std U = 2.382; $p = .017$), "*I respond to reasonable criticisms one might raise against one's viewpoints*" (Std U = 3.957; $p < .001$) and "*I provide reasons for rejecting another's claim*" (Std U = 2.588; $p = .010$).

For the SENCTDS scale, the mean range of the 21 items varies from 3.83 ("*I find that I'm easily distracted when thinking about a task*") to 5.85 ("*When faced with a decision, I seek as much information as possible*"). The standard deviation varies from 1.070 ("*When a theory,*

Co-funded by the Erasmus+ Programme of the European Union

interpretation, or conclusion is presented to me, I try to decide if there is good supporting evidence") to 1.827 ("*I find it hard to concentrate when thinking about problems*"). K-S test shows that data are equally distributed by female and male students ($p > .050$) except for items "*I often miss out on important information because I'm thinking of other things*" ($Z = 1.370$; $p = .047$), "*I like to make lists of things I need to do and thoughts I may have*" ($Z = 1.920$; $p = .001$), "*I take notes so I can organize my thoughts*" ($Z = 1.891$; $p = .002$), "*I make simple charts, diagrams or tables to help me organize large amounts of information*" ($Z = 1.598$; $p = .012$), and "*I look forward to learning challenging things*" ($Z = 1.436$; $p = .032$). The item' description is displayed in the supplementary Table 3.

Mann-Witney's U test shows no statistically significant differences in CT dispositions (SENCTDS) between female and male students ($p > 0.05$) except for nine items: "*When faced with a decision, I seek as much information as possible*" (Std U = -1.104; p=.028), "*I find it hard to concentrate when thinking about problems*" (Std U = 2.819; p = .005), "*I often miss out on important information because I'm thinking of other things*" (Std U = 2.426, p = .015), "*I like to make lists of things I need to do and thoughts I may have*" (Std U = -4.577; $p \leq .0001$), "*I take notes so I can organize my thoughts*" (Std U = -5.010; p <.0001), "*I make simple charts, diagrams or tables to help me organize large amounts of information*" (Std U = -3.557; $p \leq .0001$), "*I enjoy information that challenges me to think*" (Std U = 1.964; p = .050), "*I look forward to learning challenging things*" (Std U = 2.804; p = .005), and "*Completing difficult tasks is fun for me*" (Std U = 2.515, p = .012).

Confirmatory factor analysis (CFA) and reliability

CFA aims at confirming that the questionnaires fit the original model proposed for the original scales. For a comprehensive analysis of the structure and relations of the items in each instrument, six latent skills and a general construct were tested according to five successive models with increasing complexity.

- Model 1: One-factor model. This model tests the existence of one global factor on CT Skills or dispositions, which explains the variances of the items;

Co-funded by the Erasmus+ Programme of the European Union

- Model 2: Six-factor (non-correlated) model. This model tests the existence of six non-correlated factors that explain the variance of the set of items in each scale;

- Model 3: Six-factor (correlated) model. This model tests the existence of six correlated latent factors, each one explaining the variance of a set of items;

- Model 4: Second order factor model. This model tests how global Critical Thinking Skills or dispositions construct explains the six latent skills variance which, in turn, explain a set of items each;

- Model 5: Bi-factor model. This model tests the possibility of the scale items variances being explained by an integrated Critical Thinking Skills or Dispositions construct and by the six latent skills independently.

Table 4 shows the model fit indices for each questionnaire. In the case of the CTSAS-SF scale, the goodness-of-fit indices are satisfactory for models 3 and 4 but not for models 1, 2 and 5. As model 3 and model 4 are not nested, we guide our interpretation based on fit indices differences. The differential value of RMSEA and CFI indices between model 3 (which shows the best goodness-of-fit indices) and model 4 (which represent the original model proposed by Nair [24]) is lower than .015 and .010, respectively (ΔRMSEA = .002; ΔCFI = .003), suggesting that both models may be used to validate the internal structure of the questionnaire. As model 4 represents the original model, it will be accepted as a fitted factor structure and considered for following analysis.

Factor loadings for the CTSAS-SF scale are presented in supplementary Table 4. The loadings are significant (p<.001) and vary from 0.386 ("*I observe the facial expression people use in a given situation*") to .786 ("*I continually revise and rethink strategies to improve my thinking*"). All factor loadings are above .500 except for items «*I observe the facial expression people use in a given situation*» (.386), "*I clarify my thoughts by explaining to someone else*" (.422) and "*I confidently reject an alternative solution when it lacks evidence*" (.470).

In the case of the SENCTDS scale, the only model with a satisfactory goodness-of-fit was model 3, which was used to validate the internal structure of the questionnaire. For this scale, the

Co-funded by the Erasmus+ Programme of the European Union

factor loadings are presented in Supplementary Table 5. All loadings are significant and vary between .659 ("*Thinking is not about 'being flexible', it's about 'being right'* ") and .908 ("*I take notes so I can organize my thoughts*"). All factor loadings are above .600.

The CTSAS-SF instrument presents an excellent internal consistency (Cronbach's α =.969). Cronbach's alphas for each dimension of the scale are above .700 showing good factorial reliability (Table 4). Correlations between factors and between the factors are strong (from .750 to .965) (Table 5). All correlations are significant at p-value ≤.0001.

The SENCTDS scale presents a very good internal consistency (Cronbach's α = 0.842). Cronbach's alphas for each dimension of the scale are above .700 showing good factorial reliability (Table 4). Correlations between factors however vary from fair and moderate to strong (from .135 to .769), with exception of the non-significant correlations between *Organization* and *Attentiveness* or *Open-mindedness* (Table 5). All correlations are significant at p-value<.010.

Co-funded by the Erasmus+ Programme of the European Union

Table 4. Goodness-of-fit indices for the CTSAS-SF and SENCTDS instruments.

Scale	Models	χ^2 (df)	*p*	RMSEA [90%IC]	CFI	TLI
CTSAS-SF	Model 1: one-factor model	5159.412 (1710)	<.0001	.061 [.059-.063]	.893	.890
	Model 2: 6-factor model (non-correlated)	29275.338 (1710)	<.0001	.174 [.172-.176]	.148	.118
	Model 3: 6-factor model (correlated)	3871.243 (1695)	<.0001	.049 [.047-.051]	.933	.930
	Model 4: second-order factor model	3975.885 (1704)	<.0001	.051 [.049-.053]	.927	.924
	Model 5: Bi-factor model	18656.904 (1657)	<.0001	0.139 [.137 -.141]	.474	.439
SENCTDS	Model 1: one-factor model	4655.783 (189)	<.0001	.211 [.206-.216]	.579	.532
	Model 2: 6-factor model (non-correlated)	3828.759 (189)	<.0001	.190 [.185-.196]	.657	.618
	Model 3: 6-factor model (correlated)	447.677 (174)	<.0001	.054 [.048-.061]	.974	.969
	Model 4: second-order factor model	686.865 (183)	<.0001	.072 [.066-.078]	.952	.945
	Model 5: Bi-factor model	676.807 (165)	<.0001	.076 [.070-.082]	.952	.939

Table 5. Cronbach´s alpha reliability index for both scales, and correlations between factors plus the factors and the general CT skills construct for the CTSAS-SF scale (following the second-order factor model), and correlations between factors for the SENCTDS scale (following the six-correlated factors model).

Skills	***α***	***CT Skills***	***1***	***2***	***3***	***4***	***5***
1. Interpretation	.772	.881					
2. Analysis	.888	.925	.905				
3. Evaluation	.858	.965	.810	.934			
4. Inference	.905	.956	.806	.858	.937		
5. Explanation	.853	.907	.765	.825	.864	.868	
6. Self-regulation	.905	.851	.750	.750	.781	.841	.805

Dispositions	***α***		***1***	***2***	***3***	***4***	***5***
1. Reflection	.796	- -					
2. Attentiveness	.853	- -	.135				
3. Open-mindedness	.773	- -	.251	.396			
4. Organization	.772	- -	.431	-.020 (ns)	.077 (ns)		
5. Perseverance	.792	- -	.621	.319	.256	.355	
6. Intrinsic Goal Motivation	.842	- -	.567	.265	.321	.315	.769

Multigroup invariance for sex

To verify the factorial structure invariance of both the questionnaires across sexes, a multigroup invariance analysis was used considering the second order factor model in the case of CTSAS-SF and the six-correlated factors model in the case of SENCTDS. The WLSMV was employed as an estimation method due to the ordinal nature of the data. The initial step of the procedure was to create a baseline for both groups (female and male students) using independent CFAs for each group. Then, a CFA was applied for both groups simultaneously to test for invariance. The three invariance models, from the less restrictive (the configural model) to the most restrictive (the scalar invariance), were tested. The results are shown in Table 6.

Table 6. The goodness of fit indices for multigroup invariance by sex for the CTSAS-SF (second order factor model) and SENCTDS (six-correlated factors model).

	Baseline models	**χ^2 (df)**	***p***	***RMSEA [90%IC]***	***CFI***	***TLI***
	Female	3488.157 (1704)	<.0001	.052 [.049-.054]	.929	.926
	Male	2314.349 (1704)	<.0001	.050 [.045-.055]	.948	.946
	Invariance	**χ^2 (df)**	***p***	***RMSEA [90%IC]***	***CFI***	***TLI***
	Configural invariance	5521.460 (3390)	<.0001	.049 [.046-.051]	.939	.936
CTSAS-SF	*Metric invariance*	5490.717 (3444)	<.0001	.047 [.045-.050]	.941	.940
	Scalar invariance	5613.987 (3732)	<.0001	.044 [.041-.046]	.946	.949
	Model comparison	**χ^2 (df)**	***p***	***ΔRMSEA***	***ΔCFI***	
	Metric vs. Configural	45.988 (54)	.773	.002	.002	
	Scalar vs. Configural	370.658 (342)	.137	.005	.007	
	Scalar vs. Metric	328.786 (288)	.049	.003	.005	
	Baseline models	**χ^2 (df)**	***p***	***RMSEA [90%IC]***	***CFI***	***TLI***
	Female	352.859 (174)	<.0001	.051 [.044-.059]	.977	.973
	Male	313.264 (174)	<.0001	.075 [.062-.088]	.953	.943
	Invariance	**χ^2 (df)**	***p***	***RMSEA [90%IC]***	***CFI***	***TLI***
	Configural invariance	660.692 (348)	<.0001	.058 [.051-.065]	.970	.964
SENCTDS	*Metric invariance*	663.475 (363)	<.0001	.056 [.049-.063]	.971	.966
	Scalar invariance	754.103 (460)	<.0001	.049 [.043-.055]	.972	.974
	Model comparison	**χ^2 (df)**	***p***	***ΔRMSEA***	***ΔCFI***	
	Metric vs. Configural	14.683 (15)	.4745	.002	.001	
	Scalar vs. Configural	123.360 (112)	.2180	.009	.002	
	Scalar vs. Metric	110.600 (97)	.1632	.007	.001	

Co-funded by the Erasmus+ Programme of the European Union

Based on the goodness of fit values of the different invariance models tested (configural, metric and scalar), in the CTSAS-SF instrument, the stability of the factor structure in both sexes is confirmed. The difference (Δ) in CFI and RMSEA values between models is less than .015 and .010, respectively, revealing the invariance of the factorial structure, the invariance of factor loadings and the invariance of the item intercepts when comparing female and male students. Similarly, the SENCTDS instrument presents a good stability of the factor structure in both the sexes. The difference in CFI and RMSEA values between models is less than .010 and .015 [32, 33], respectively, thus confirming the invariance of the factorial structure, the factor loadings and of item intercepts when comparing female and male students. Once the instrument invariance confirmed, the structural invariance related to the populational heterogeneity as well as the latent mean invariance were tested. Structural invariance tests whether the covariance level between factors is the same for both groups. Latent mean invariance assesses whether the latent means are equal in both groups.

Table 7 displays the results from the structural invariance in both groups in CTSAS-SF and SENCTDS. Wald's test shows a significant difference between factor correlations of the female and male models in CTSAS-SF (Wald = 6.507; df = 1; p = .011) but not in SENCTDS (Wald=.316; df=1; p=.5742). In CTSAS-SF, the factor covariances are significantly higher on the male model than in the female model suggesting some population heterogeneity. In contrast, in SENCTDS, the factor covariances are similar between male and females, suggesting a more uniform behaviour between sexes.

Co-funded by the Erasmus+ Programme of the European Union

Table 7. Factor covariances by sex for the CTSAS-SF (CT skills) and SENCTDS (CT Dispositions).

Skills	***Interpretation***		***Analysis***		***Evaluation***		***Inference***		***Explanation***	
	F	***M***	***F***	***M***	***F***	***M***	***F***	***M***	***F***	***M***
Analysis	.888	.941								
Evaluation	.760	.900	.922	.955						
Inference	.759	.890	.838	.902	.924	.956				
Explanation	.739	.849	.816	.877	.850	.907	.856	.925		
Self-regulation	.720	.808	.738	.780	.759	.825	.805	.907	.782	.885

Dispositions	***Reflection***		***Attentiveness***		***Open-mindedness***		***Organization***		***Perseverance***	
	F	***M***	***F***	***M***	***F***	***M***	***F***	***M***	***F***	***M***
Attentiveness	.119**	.220***								
Open-mindedness	.242	.286	.428	.348						
Organization	.561	.528	-.003 (ns)	.017 (ns)	.151*	-.144 (ns)				
Perseverance	.559	.553	.246	.495	.324	.109 (ns)	.401	.281		
Intrinsic Goal Motivation	.572	.589	.232	.309	.387	.196***	.382	.338	.767	.807

F = Female students, M = Male students. All correlations are significant at p-level < .001 for the CTSAS-SF scale. In SENCTDS, all the correlations are significant at p-level level <.001, except for: * p=.003; **p=.024; ***=.005; ns= non-significant

At the means invariance analysis, female students are the baseline group with a latent mean equal to zero. The mean comparisons for both the CTSAS-SF and the SENCTDS are presented in Table 8. There are non-significant differences in factor means between females and males in CTSAS-SF but in SENCTDS, males presented significantly higher average scores in *Attentiveness* and *Intrinsic Goal Motivation*, and a significantly lower average score in *Organization*.

Table 8. Latent means differences between female and male.

Skills		***ΔMeans***	***SE***	***Est/SE***	***p***
CTSAS-SF	*Interpretation*	-.014	.106	-.129	.897
	Analysis	.023	.096	.244	.807
	Evaluation	.071	.096	.736	.462
	Inference	-.051	.099	-.512	.608
	Explanation	.177	.097	1.832	.067
	Self-regulation	-.005	.098	-.046	.963
Dispositions					
SENCTDS	*Reflection*	-.197	.118	-1.676	.094
	Attentiveness	.206	.103	1.994	**.046**
	Open-mindedness	-.120	.108	-1.111	.266
	Organization	-.511	.110	-4.647	**≤.0001**
	Perseverance	.024	.108	0.218	.826
	Intrinsic goal motivation	.264	.105	2.507	**.012**

Multigroup invariance by country

The study of the factorial structure invariance across the countries (disciplines) was only possible for the three countries with larger number of respondents (n=445), namely Greece (n=177), Portugal (n=156), and Romania (n=112). For this analysis, the baseline for all the groups used independent CFAs for each group; the CFA was applied simultaneously to the three groups to test for invariance. The results for the three invariance models tested are shown in Table 9.

Co-funded by the Erasmus+ Programme of the European Union

Table 9. The goodness of fit indices for multigroup invariance by country (Discipline) for the CTSAS-SF (second order factor model) and SENCTDS (six-correlated factors model).

	Baseline models	***χ² (df)***	***p***	***RMSEA [90%IC]***	***CFI***	***TLI***
	Overall model	3639.787 (1695)	<.0001	.051 [.049-.053]	.929	.925
	Portugal	2528.042 (1695)	<.0001	.056 [.052-.061]	.902	.898
	Greece	2659.082 (1695)	<.0001	.057 [.056-.061]	.908	.904
	Romania	2117.048 (1695)	<.0001	.047 [.040-.053]	.965	.963
	Invariance	***χ² (df)***	***p***	***RMSEA [90%IC]***	***CFI***	***TLI***
CTSAS-SF	*Configural invariance*	7427.003 (5085)	<.0001	.056 [.053-.058]	.929	.926
	Metric invariance	--	--	--	--	--
	Scalar invariance	7922.401 (5769)	<.0001	.050 [.047-.053]	.935	.940
	Model comparison	***χ² (df)***	***p***	***ΔRMSEA***	***ΔCFI***	
	Metric vs. Configural	--	--	--	--	
	Scalar vs. Configural	849.399 (684)	<.0001	.006	.006	
	Scalar vs. Metric	--	--	--	--	
	Baseline models	***χ² (df)***	***p***	***RMSEA [90%IC]***	***CFI***	***TLI***
	Overall model	528.715 (195)	<.0001	.062 [.056-.068]	.963	.957
	Portugal	316.839 (174)	<.0001	.073 [.060-.085]	.959	.950
	Greece	332.035 (174)	<.0001	.072 [.060-.083]	.955	.946
	Romania	257.725 (174)	<.0001	.066 [.048-.082]	.973	.967
	Invariance	***χ² (df)***	***p***	***RMSEA [90%IC]***	***CFI***	***TLI***
SENCTDS	*Configural invariance*	905.093 (522)	<.0001	.070 [.063-.078]	.962	.955
	Metric invariance	--	--	--	--	--
	Scalar invariance	1192.983 (744)	<.0001	.064[.057-.070]	.956	.963
	Model comparison	***χ² (df)***	***p***	***ΔRMSEA***	***ΔCFI***	
	Metric vs. Configural	--	--	--	--	
	Scalar vs. Configural	358.373 (222)	<.0001	.006	.006	
	Scalar vs. Metric	--	--	--	--	

The results related to the scalar invariance for the CTSAS-SF scale (RMSEA = .050; CFI = .935; TLI = .940) are better than those found for the overall model (RMSEA = .051; CFI = .929; TLI=.925). Moreover, the difference between the configurational and the scalar invariances is less than .015 for RMSEA values, and less than .01 for the CFI values, thus confirming the invariance between countries/disciplines. Differences were observed when comparing the average values among the three countries, for all the six latent variables (skills). The exception was found between Greece and Portugal for the skills *Interpretation* and *Self-regulation*. In the significant comparisons, the average values are higher for the group of Greek and Romanian students, compared to the Portuguese. The factorial comparison between countries (Portugal providing the baseline, with average 0 and variance 1) is provided in Table 10.

Co-funded by the Erasmus+ Programme of the European Union

Table 10. Latent means differences between Greece, Romania, and Portugal, using the latter as baseline with the average=0 and variance=1.

Skills		***Factor Means***		***SE***		***Est/SE***		***p***	
		Greece	***Romania***	***Greece***	***Romania***	***Greece***	***Romania***	***Greece***	***Romania***
CTSAS-SF	*Interpretation*	-.001	.398	.108	.113	-.010	3.514	.992	<.0001
	Analysis	.232	.547	.108	.114	2.158	4.792	.031	<.0001
	Evaluation	.226	.633	.112	.124	2.020	5.124	.043	<.0001
	Inference	.468	.609	.108	.116	4.340	5.255	<.0001	<.0001
	Explanation	.257	.516	.109	.115	2.362	4.475	.018	<.0001
	Self-regulation	.076	.335	.105	.116	.720	2.894	.471	.004
Dispositions									
SENCTDS	*Reflection*	.475	.371	.124	.129	3.820	2.881	<.0001	.004
	Attentiveness	-.082	.294	.104	.132	-.791	2.232	.429	.026
	Open-mindedness	-1.054	-.538	.154	.155	-6.834	-3.476	<.0001	.001
	Organization	.044	.184	.117	.148	.376	1.241	.707	.214
	Perseverance	-.024	.119	.141	.152	-.169	.783	.866	.434
	Intrinsic goal motivation	-.372	.103	.130	.138	-2.861	.745	.004	.456

In contrast, the multigroup invariance computed for SENCTDS scale shows slightly worse results for the countries when compared to the overall model. The scalar invariance indices (RMSEA=.064; CFI = .956; TLI = .963) and the small differences when comparing the scalar vs. configural invariances, however, confirm that the factorial structure is maintained stable in the country groups analysed. Nonetheless, this assumption should be taken with precaution, as the comparison of the factorial averages between Greece, Romania, and Portugal (Table 11) shows that in the multigroup invariance only two of the six latent factors do not present differences.

The estimated correlations show that the averages for CT skills are higher in Greece and Romania compared to Portugal. Regarding the CT dispositions, the estimated correlations evidenced higher averages in Greece and Romania than in Portugal, but the inverse (Portugal presenting the highest averages) was noticed for *Open-mindedness*. Romania scored better than Portugal in *Attentiveness*; while Greece scored lower than Portugal in *Intrinsic goal motivation*.

It was not possible to compute the Wald index, but Table 11 presents the correlations between countries for CT skills (CTSAS-SF) and dispositions (SENCTDS). In general, the factorial correlation between countries remains; the exceptions to these trends are signalled in bold in Table 11.

Table 11. Factor covariances by country for the CTSAS-SF (Skills) and SENCTDS (Dispositions)

CTSAS-SF	*Country*	*Analysis*	*Evaluation*	*Inference*	*Explanation*	*Self-regulation*
Interpretation	***POR***	*.910***	*.799***	.762**	.715**	.759**
	GRE	*.906***	*.785***	.790**	.741**	.673**
	ROM	.872**	.801**	.805**	.754**	.741**
Analysis	***POR***		.924**	.814**	.786**	.679**
	GRE		.955**	.854**	.810**	.710**
	ROM		.892**	.882**	.848**	.783**

Co-funded by the Erasmus+ Programme of the European Union

CTSAS-SF	Country	Analysis	Evaluation	Inference	Explanation	Self-regulation
Evaluation	POR			.886**	.763**	.726**
	GRE			.954**	.903**	.735**
	ROM			.926**	.882**	.840**
Inference	POR				.842**	.783**
	GRE				.880**	.804**
	ROM				.880**	.885**
Explanation	POR					.774**
	GRE					.778**
	ROM					.868**

SENCTDS	Country	Attentiveness	Open-mindedness	Organization	Perseverance	Intrinsic Goal Motivation
Reflection	POR	.395**	.495**	.469**	.467**	.532**
	GRE	.289**	.443**	.421**	.566**	.545**
	ROM	.416**	.464**	.579**	.608**	.596**
Attentiveness	POR		.250**	.072(ns)	.436**	.293**
	GRE		.559**	-.103(ns)	.114(ns)	.074(ns)
	ROM		.388**	-.013(ns)	.516**	.370**
Open-mindedness	POR			.217**	.500**	.460**
	GRE			-.113(ns)	.203**	.243**
	ROM			.296**	.297**	.287**
Organization	POR				.395**	.217**
	GRE				.263**	.333**
	ROM				.568**	.532**
Perseverance	POR					.758**
	GRE					.656**
	ROM					.834**

POR- Portugal; GRE – Greece; ROM – Romania. All correlations are significant at p-level <0.0001

4. Discussion of the instrument' validation process

The instrument used to monitor the changes in CT skills and dispositions results from the junction of two distinct self-report questionnaires, one targeting the CT skills (adapted CTSAS

Co-funded by the Erasmus+ Programme of the European Union

or CTSAS-SF) and the other aiming to assess the CT dispositions (SENCTDS), which was used in its original form. The CTSAS short form that the consortium elaborated, by reducing the time to complete the questionnaire, allowed for the filling of both questionnaires in less than 60 minutes. This time-lapse was deemed comfortable for students.

For the validation of the final questionnaire, the first set of responses submitted by the students enrolled in the pilot courses (pre-test moment) was used. The questionnaire was applied at the beginning of the pilot CTBACs implemented in the first and second semester of 2021/2022, to a total of 531 students in the five Higher Education Institutions. Responses were collected during the first week of the Course, before the beginning of the learning interventions.

The age of the respondents ranged between 19 and 58 years old (mean=23,5), with 87% of the students under the age of 31. Considering that in Europe, in general, women are in the majority in HEI, especially in areas such as Health Sciences, Arts, Humanities, Social Sciences and Education, it was not surprising that a higher percentage of the respondents were females (75%; 389 females vs.142 males). The small percentage of students (6,4%) in the fields of Informatics and Technology, typically more frequented by males, was not enough to reduce the percentage difference between respondents of each sex.

CTSAS validation

During the development of the Think4jobs project it became clear that the instrument to be applied should have a reduced size (regarding the number of items), to make it possible to be completed in a lapse of time comfortable for the students. CTSAS questionnaire, authored by Nair [24] had a total of 115 items (ranging from 0=never to 6=always), which corresponded to a completion time of 50 minutes. With the process of reducing the number of items (elimination of redundant items and elimination of items with loading weights below .500) the CTSAS SF questionnaire had an average response time of less than 30 minutes, keeping

Co-funded by the Erasmus+ Programme of the European Union

the original six dimensions (*Interpretation*, *Analysis*, *Evaluation*, *Inference*, *Explanation* and *Self-regulation*), with only 60 items.

The CTSAS-SF validation process, with confirmatory factor analysis, resulted in the obtainment of two models with equivalent satisfactory goodness-of-fit indices. Model 4, the second order factor model (RMSEA=.051; TLI=.924; CFI=.927) had a chi-square/df ratio of 2.33 and Model 3 the correlated six-factor model had a chi-square/df ratio of 2.28 (RMSEA=.049; TLI=.930; CFI=.933), which confirms that both models have a very good overall fitness, internal consistency, and reliability. The Cronbach alpha of the overall instrument was excellent (α = .969) [35, 36].

The second order factor model (Model 4) was considered having a better fit with the idea that critical thinking is a complex multidimensional and multileveled construct [37, 38]; also, this idea is further supported by the fact that the model that tested the hypothesis that all the 60 items are explained by one factor (Model 1) and the bi-factorial model (Model 5) had not an adequate fit to the data. The factor analysis supported a six-factor correlated structure: *Interpretation* (9 items; α=.772), *Analysis* (11 items; α=.888), *Evaluation* (7 items; α=.858), *Inference* (13 items; α=.905), *Explanation* (10 items; α=.853) and *Self-regulation* (10 items; α=.905).

When we refer to the concept of CT, we have to take into account its higher-order nature [39], and bear in mind that “differences in a learner's history may mean that a situation that requires higher order thinking by one person may need only lower order thinking by another person. Furthermore, in the classroom, the teaching of basic and higher order thinking skills are likely to be interwoven.” [39].

An important aspect that still deserves some attention is the fact that four items (items 4, 6, 8, and 39), had a factorial load below 0.500. This situation was also verified in Nair’s study, and the team of researchers decided to keep the four items, as we considered that the

Co-funded by the Erasmus+ Programme of the European Union

substantive content of each of these items was important for the characterization of essential aspects of CT (namely in Interpretation and Inference domains).

The correlations found between skills and between the skills and the integrated CTSAS-SF score were strong and positive, supporting the existence of a good item-related validity and confirming the very good internal consistency and reliability. Sex did not affect the data distribution, except in four items (items 4, 42, 47 and 50). Moreover, the CTSAS-SF maintained its factorial structure invariance across sexes, supporting its reliability for both sexes and countries.

With good model-data fit measures, the CTSAS-SF has a very good validity and reliability and despite the reduction in the number of items, the scale maintained its stability. It demonstrates a high potential for its use in research requiring the assessment of CT in higher education students and showed to produce good results even when applied in a multinational context to students in five very different higher education programmes.

SENCTDS validation

The SENCTDS validation process, with confirmatory factor analysis, identified only one model presenting a satisfactory goodness-of-fit indices – the model 3, of six- correlated factors (RMSEA=.054; TLI=.974; CFI=.969) with a chi-square/df ratio of 2.57. These values evidence a very good overall fitness, internal consistency, and reliability of the model. The SENCTDS instrument further presented a high Cronbach alpha (α = .842), suggestive of the strong internal consistency of the instrument [40]. Moreover, the individual dimensions of the CT dispositions assessed with SENCTDS presented acceptable-to-good Cronbach's alpha values [35, 36, 40]: *Reflection* (3 items; α=.796), *Attentiveness* (4 items; α=.853), *Open-Mindedness* (4 items; α=.773), *Organization* (3 items; α=.772), *Perseverance* (3 items; α=.792) and *Intrinsic Goal Motivation* (4 items; α=.842). These coefficients support the conclusion that the constructs measure the intended dimensions, suggesting that each one of the six scales can

Co-funded by the Erasmus+ Programme of the European Union

be used independently whenever a particular dimension needs to be assessed in separate [41].

In SENCTDS, even if a rather uniform behaviour has been suggested between females and males according to the factorial covariances, the invariance analysis evidenced that males scored higher in *Attentiveness* and *Intrinsic Goal Motivation* but lower in *Organization* when compared to females. The invariance analysis for country/discipline confirmed that the factorial structure was maintained between countries, even if only two of the six dispositions' factors fail to present differences in the factorial means among countries. Even though this behaviour may hint at existing differences between countries in regards to the scale behaviour, we can not discard the effects of the unbalanced country representativeness among respondents [42].

Co-funded by the Erasmus+ Programme of the European Union

PART II – CROSS DISCIPLINARY ANALYSIS OF CTBACS IMPLEMENTATION

The evaluation of teaching effectiveness is crucial, particularly when new instructional strategies are designed to target students' development in a particular area (whether cognitive or attitudinal development) or are implemented in a course or across disciplines. The evaluation of the results for such interventions drives and supports curricular changes and should not be overlooked.

In the absence of a standard instrument for evaluating learning effectiveness, teachers and institutions often resource to a pre-test/post-test evaluation of the targeted learning goals in a course (short-term learning and longitudinal assessment), or in the context of an educational programme and the quality of the provided training (long term learning and transversal assessment) [43]. The pre-test/post-test approach allows to establish the baseline competence of students at the beginning of the instruction and measures improvements at the end of the instruction [43]. Even though this pre-test/post-test approach is often used, in many cases it targets the assessment of the cognitive knowledge, and less frequently the soft skills or dispositions trained with the students.

Besides, the regular use of the pre-test/post-test approach across graduate programs, as a mean for a sound monitoring strategy, would further allow to track the students' knowledge and competencies and methodically adjust them to mitigate arising gaps derived from the constant technological development or the critical challenges found in most professions.

Nonetheless, the literature available on the assessment of the implementation of new pedagogical interventions suggest the existence of some difficulties in the efficacy assessment, namely in the context of multiple instructional interventions (in form and time length). Moreover, regarding the poor categorization of the instructional intervention, the absence of a pre-test to establish the baseline, the use of non-validated tools, and the collection of student' perceptions of gains, or the small number of students in the study. Also,

Co-funded by the Erasmus+ Programme of the European Union

most studies are quasi-experimental, as it is difficult to constitute a control group within the same course [17]. A similar problem was faced in this study, with only two countries being able to enrol a control group.

In this section the Report aims at evaluating any putative short-term changes in students' CT skills and dispositions after the delivery of the twelve CT-blended apprenticeship curricula presented in IO3 [1]. In this evaluation, the focus was established in the enhancement of CT skills and dispositions within a course, and not in the acquisition of cognitive knowledge specific to that course, even though the CT dimensions have been grounded in the specific context of a course or discipline. To analyse those changes, the CTSAS-SF and SENCTDS scales (validated for this population) were used before and after the pedagogical interventions.

The data gathered during this process are analysed according to the changes recorded from the comparison of the students' scores at the beginning and the end of the piloting courses (cross-country comparisons) but also within each country, as it is difficult to make sensible comparisons between countries due to the existing confounding variables, namely:

- The differences in the disciplines represented in the study (one discipline – one country);
- The differences in students, issued from the individual (cultural or religious, among others) and disciplinary backgrounds as well as the differences in their engagement with the activities;
- The differences in the interventions implemented;
- The use of different instructors in diverse courses, which may affect the interpersonal relationships and the way learning is achieved.

Co-funded by the Erasmus+ Programme of the European Union

1. Methodology

The Project proposed to measure skills and dispositions associated to CT as a measure of the success of the implemented CTBACs in pilot courses in Business Informatics (Germany), Teacher Education (Greece), English as Foreign Language (Lithuania), Veterinary Medicine (Portugal) and Business and Economics (Romania). In the submitted proposal, the use of the CT skills and dispositions scoring instruments was planned in three moments during each course duration. An experimental approach was also foreseen, by creating an experimental and control groups.

The CTSAS-SF and SENCTDS instruments were used to establish the changes in CT skills and dispositions, respectively. However, we faced some constraints to fulfil the proposed plan. For one, in the period scheduled for CTBACs implementation, the pilot courses were offered only once in most countries/disciplines (German, Lithuania; Portugal and Romania). On the other hand, the students claimed that creating a control group within the same curricular unit might lead to inequitableness in the learning quality, as they saw the differences in learning interventions as a discrimination, with the experimental group receiving a putative advantage. Consequently, all the students registered in the piloting course were submitted to the proposed interventions. For all the countries except Greece and Portugal, the study evolved as a quasi-experimental situation.

Another constraint faced was related to the loss of students across the three moments of the application of the CT scoring instruments. Some of them failed to complete one of the questionnaires. Consequently, at the end of the study, very few students presented the three completed questionnaires requested; often the intermediate one was amiss. Therefore, the partnership decided to use only the data gathered at the pre-test (moment zero of the course) and the post-test (the one filled at the end of the course) to evaluate the success of the pedagogical interventions.

Co-funded by the Erasmus+ Programme of the European Union

The links to the merged CTSAS-SF and SENCTDS questionnaires were shared with the students via the Moodle page for each course, as explained before. Besides the English version (used by the Lithuanian partners), four translations of the instruments were provided in German, Greek, Portuguese and Romanian (see section I for details).

Participants

The current study used a non-randomized, convenience sample from the students enrolled in the piloting courses where the CTBACs were implemented. Of the 609 students enrolled in pilot activities, only 87.2% responded to questionnaires. From the 531 students that responded to the pre-test, at the beginning of the course, only 63.1% of the students responded to the post-test questionnaire completed at the end of the course. The paired questionnaires reached a representativeness of 55% of the population involved in the project. Consequently, the population used to assess the success of the pilot interventions is composed of 335 students (table 12); 258 (77%) were female and 77 (23%) were male. The unbalanced distribution of sexes followed the pattern described in part I of this report. The average age was 23.62 years (sd = 7.50; range 18 –58).

The experimental group comprised 286 students (85.4%) while 49 students (14.6%) constituted the control group. The unbalanced representativeness of the control group advises the reader to consider any interpretation of the results with caution. Students represented five countries; most of them were Portuguese (32.5%), Greek (30.7%) and Romanian (24,2%). Students from Germany (6.6%) and Lithuania (6.0%) represented only 12.6% of the participants (Table 12). The control group was composed by Greek (n= 40) and Portuguese (n=9) students.

Table 12. Distribution of the students' population per discipline and course (CTBACs and Control) and their representativeness for the database of paired questionnaires.

Co-funded by the Erasmus+ Programme of the European Union

Country	Programme/Discipline	Courses	No. students that responded to		% of the total paired questionnaires
			Pre-test	Post-test	
Germany	Business Informatics	Design patterns	11	8	2,4
		Innovation Management	10	4	1.2
		Economic Aspects of Industrial Digitalization	10	8	2.4
		Scientific seminar	10	2	.6
Greece	Teacher Education	Teaching Biological Concepts	46	22	6.6
		Teaching Science Education	111	74	22.1
		Teaching of the Study of the Environment	20	7	2.1
Lithuania	International Relations and Political Science	English as Foreign Language	52	20	6.0
Portugal	Veterinary Medicine Integrated Masters	Imaging	55	36	10.7
		Deontology	52	42	12.5
		Gynecology. Andrology and Obstetrics	41	22	6.6
		Curricular Traineeship	12	9	2.7
Romania	Business and Economics	Business Communication	35	31	9.3
		Pedagogy and Didactics of Financial Accounting	40	32	9.6
		Virtual Learning Environments in Economics	26	18	5.4
Total			**531**	**335**	**100**

Characterization of the experimental group

The experimental group integrated 286 students, from which 76.2% (n = 218) were female and 23.8% (n = 68) male. The average age was 23.88 years (sd = 7.62; range 18 – 54). 35% (n = 100) of students were from Portugal, followed by 28.3% (n = 81) from Romania and 22.0% (n = 63) from Greece. Students from Germany and Lithuania represent 7.7% (n = 22) and 7.0% (n = 20) of the students, respectively. Student's distribution per discipline and course is shown in Table 13, and age distribution in Table 14.

Table 13. Distribution of the students' experimental group per discipline and course

Programme/Discipline	Courses	n	%
Business Informatics (n=22)	Design patterns	8	2.8
	Innovation Management	4	1.4
	Economic Aspects of Industrial Digitalization	8	2.8
	Scientific seminar	2	.7
Teacher Education (n=63)	Teaching Biological concepts	22	7.7
	Teaching Science Education	34	11.9

Co-funded by the Erasmus+ Programme of the European Union

Programme/Discipline	Courses	n	%
Teacher Education (n=63)	Teaching of the Study of the Environment	7	2.4
International Relations and Political Science (n=20)	English as Foreign Language	20	7.0
Veterinary Medicine Integrated Masters (n=100)	Imaging	36	12.6
	Deontology	42	14.7
	Gynecology. Andrology and Obstetrics	22	7.7
Business and Economics (82)	Business Communication	31	10.8
	Pedagogy and Didactics of Financial Accounting	32	11.2
	Virtual Learning Environments in Economics	18	6.3
Total		**286**	**100.0**

Characterization of the control group

The control group was composed of 49 respondents, 40 from Greece (Teacher Education discipline – Teaching Science Education Course) and nine from Portugal (Veterinary Medicine discipline– Curricular Traineeship Course). The mean age of the respondents was 22.10 years (sd =6.61; range 19 –58). Forty students were female and nine were male.

Table 14. Age distribution in the experimental and control groups

		Experimental Group		*Control Group*	
		N	**Percentage**	***N***	**Percentage**
Age (years)	18	7	2,45	0	0
	19	33	11,54	16	32,7
	20	76	26,57	16	32,7
	21	52	18,18	5	10,2
	22	20	6,99	0	0
	23	33	11,54	3	6,1
	24	7	2,45	3	6,1
	25	8	2,80	1	2,0
	26	6	2,10	1	2,0
	27	4	1,40	0	0
	28	0	0,00	1	2,0
	29	4	1,40	0	0

Co-funded by the Erasmus+ Programme of the European Union

		Experimental Group		Control Group	
		N	Percentage	N	Percentage
Age (years)	30-35	7	2,45	0	0
	36-40	8	2,80	2	4,1
	41-45	8	2,80	0	0
	46-50	8	2,80	0	0
	51-55	5	1,75	0	0
	56-60	0	0,00	1	2,0
	Total	**286**	**100**	**49**	**100,0**

Data analysis

In the CTSAS-SF scale, the skills (dimensions) are composed of two to three subskills (subdimensions), which in turn are compounded by a different number of items. To compute the score of each subskill the average of scores obtained in the items composing the corresponding subdimension were used. To compute the score for a particular skill, the sum of the values obtained for each corresponding subdimension was used. The range values for skills and subskills are provided in Table 15.

Table 15. Range of scores for each dimension and subdimension of the SCTSAS-SF and SENCTDS scales.

Skills	Skill score range	Subskills	Items	Range of subskills score (Mean)
Interpretation	0 – 18	*Categorization*	1 - 3	0 – 6
		Clarifying meaning	6 - 9	0 – 6
		Decoding significance	4, 5	0 – 6
Analysis	0 – 18	*Detecting arguments*	15, 16	0 – 6
		Analysing arguments	17 - 20	0 – 6
		Examining ideas	10 - 14	0 – 6
Evaluation	0 – 12	*Assessing claim*	21, 22	0 – 6
		Assessing arguments	23 - 27	0 – 6

Co-funded by the Erasmus+ Programme of the European Union

Skills	Skill score range	Subskills	Items	Range of subskills score (Mean)
Inference	0 – 18	*Drawing conclusions*	36 - 40	0 – 6
		Conjecturing alternatives	31 - 35	0 – 6
		Querying evidence	28 - 30	0 – 6
Explanation	0 – 18	*Stating results*	41, 42	0 – 6
		Justifying procedures	43, 44	0 – 6
		Presenting arguments	45 -50	0 – 6
Self-regulation	0 – 12	*Self-examining*	51 - 57	0 – 6
		Self-correction	58 - 60	0 – 6
Integrated score	0 – 96		1 - 60	

Dispositions	Items	Range of dispositions' score (mean)
Reflection	1 - 3	1 – 7
Attentiveness	4 - 7 (reverted)	1 – 7
Open-mindedness	8 - 11 (reverted)	1 – 7
Organization	12 - 14	1 – 7
Perseverance	15 - 17	1 – 7
Intrinsic goal motivation	18 - 21	1 – 7
Integrated score	1 - 21	6 – 42

Respecting the SENCTDS scale, before calculating the mean score for each dimension, items 4 to 11 were reversed. The score for each dimension (disposition) of the scale was computed as a mean. The range values for the dimensions of the SENCTDS instrument are provided in Table 15. The integrated scores for skills and dispositions represent the sum of all the dimensions of the scale. After calculating the results for each dimension and subdimension, the statistical analysis was performed.

Statistical analysis

Demographic data (sex, age and country) were analysed using descriptive statistics. T-test for independent samples, One-way ANOVA and Pearson correlation were used to test baseline differences for sex, course, and age, respectively.

Co-funded by the Erasmus+ Programme of the European Union

The Paired t-test was used to evaluate the general changes in the CT skills and dispositions' scores measured at two different time points (before and after the last pedagogical intervention). One-way ANOVA was applied to test country differences on changes before and after the pedagogical intervention.

To analyse the effect of different strategies by country, a GLM-Univariate Ancova was used with the score after the intervention as a dependent variable, the different strategies as independent variable and the score before the intervention as a covariate.

All analyses were developed using IBM SPSS Statistics 26. All tests were applied with the statistical significance level set to p≤.05 and the confidence interval set at 95%.

2. Results

Testing for baseline differences in the experimental group

The existence of baseline differences between sexes and countries for the experimental group has been tested for both CT skills and dispositions. No significant differences were found in the pre-test scores between sexes for the CT skills' integrated score, nor for the individual skills and subskills. Nonetheless, a residual significant difference was found in the subskill "*Explanation: presenting arguments*" (p=.055), where male students presented higher mean compared with the female students (3.85 vs 3.59).

Significant differences were found between males and females in the dimensions *Reflection* (p = .022), *Open-Mindedness* (p = .018) and *Organization* (p ≤ .0001), as well as in the dispositions' integrated score (p = .013). In all the four situations, females presented higher average scores compared with males (*Reflection*: 5.81 vs 5.52; *Open-Mindedness*: 5.52 vs 5.13; *Organization*: 5.13 vs 4.25; *Dispositions integrated score*: 31.24 vs 29.75).

Co-funded by the Erasmus+ Programme of the European Union

Significant differences were found between countries in all the CT skills, subskills, and the integrated score, with exception of the subskills: *Interpretation-Clarifying meaning* (p = .396), *Explanation-Justifying Procedures* (p = .052) and *Explanation-presenting arguments* (p=.131). In general, Romanian students present the higher average scores in all the skills and subskills as well as in the integrated score. Lithuanian students presented the lowest average scores. Nevertheless, the small sample of Lithuanian and German students recommends caution when interpreting these results.

Regarding the CT dispositions scale, no significant differences were found in the average scores between the five countries with exception for the *Open-Mindedness dimension*, whose averages are presented in Table 16.

Table 16. Average scores per country in the *Open-Mindedness* dimension

Country	n	Mean	SD
Portugal	100	5.91	1.08
Romania	81	5.18	1.28
Greece	63	5.06	1.19
Lithuania	20	5.51	.80
Germany	22	5.13	1.18
Total	286	5.43	1.21

Relationship between age and CT Skills and Dispositions

A statistical positive relation was found between students' age and the CT skills' integrated score (r = .193; p = .001) and each CT skill with exception of *Evaluation* (r = .097; p = .102). A positive association was also found between age and CT subskills, except for the subskills identified in table 17.

Co-funded by the Erasmus+ Programme of the European Union

A positive significant association with age was found in 3 of the 6 dimensions of the disposition scale and in the integrated score; the exceptions were *Attentiveness* (r = .035; p = .558) *Open Mindedness* (r = -.04; p = .948), and *Organization* (r = .106; p = .073) (Table 17).

Table 17. Baseline CT Skills and subskills and Dispositions in the experimental group not affected by students' age

Skill and Subskill	**r**	**p**
Interpretation_Decoding Significance	.083	.162
Evaluation_Assessing Claim	.061	.305
Explanation_Justifying Procedures	.097	.101
Explanation_Presenting Arguments	.076	.198
Self-Regulation_Self-Correction	.096	.107
Dispositions	**r**	**p**
Attentiveness	.035	.558
Open Mindedness	-.004	.948
Organization	.106	.073

Simple overall pre- and post-test comparison for the experimental group

The overall comparison of pre-test/post-test scores (without considering the initial differences, reported earlier) confirm a gain in CT skills in the students' experimental group (table 16). However, the gain for dispositions is not so patent; only the dimension *Organization* showed a significant improvement. As shown in Table 18, the *Open-Mindedness* disposition significantly decreased between the pre- and post-test scores.

Table 18. Comparison of the CT Skills and Dispositions in the group of students enrolled in the *CTBACs in the pre-test and post-test measurements*

Co-funded by the Erasmus+ Programme of the European Union

SKILLS	Mean	SD	Mean difference	t	p
POST_ Interpretation	13.67	2.20	1.01	7.357	≤.0001
PRE_ Interpretation	12.66	2.31			
POST_ Evaluation	8.85	1.72	.73	7.134	≤.0001
PRE_ Evaluation	8.13	1.85			
POST_ Analysis	13.08	2.52	1.12	7.650	≤.0001
PRE_ Analysis	11.96	2.74			
POST_ Inference	13.42	2.51	.95	6.779	≤.0001
PRE _Inference	12.47	2.63			
POST_ Explanation	12.88	2.50	1.42	9.688	≤.0001
PRE_ Explanation	11.46	2.66			
POST_ Self-Regulation	9.17	1.77	.61	6.151	≤.0001
PRE_ Self-Regulation	8.57	1.93			
POST_ INTEGRATED SCORE	71.09	11.79	5.85	9.705	≤.0001
PRE_INTEGRATED SCORE	65.24	12.02			
DISPOSITIONS	**Mean**	**SD**	**Mean difference**	**t**	**p**
POST_ Reflection	5.84	.88	.10	1.766	.079
PRE_ Reflection	5.74	.89			
POST_ Attentiveness	3.94	1.46	-.06	-.805	.422
PRE_ Attentiveness	4.00	1.42			
POST_ Open-Mindedness	5.23	1.43	-.20	-2.636	.009
PRE_ Open-Mindedness	5.43	1.21			
POST_ Organization	5.10	1.32	.18	2.568	.011
PRE_ Organization	4.92	1.41			
POST_ Perseverance	5.47	1.17	.07	1.182	.238
PRE_ Perseverance	5.40	1.12			
POST_ Intrinsic Goal Motivation	5.49	1.08	.10	1.712	.088
PRE_ Intrinsic Goal Motivation	5.39	1.07			
POST_ INTEGRATED SCORE	31.07	4.65	.18	.854	.394
PRE_ INTEGRATED SCORE	30.88	4.33			

Co-funded by the Erasmus+ Programme of the European Union

Cross disciplinary comparisons for the experimental group

In this Project, each Country represented a particular discipline. Therefore, for the purpose of this analysis, the terms “country” and “discipline” overlap and are used interchangeably.

The countries significantly differed in the mean difference between the pre-test/post-test scores for the CT skills integrated score ($F = 2.993$; $df = 4$; $p = .019$), but not for the CT dispositions’ integrated scores ($F = 1.808$; $df = 4$; $p = .127$). Regarding the CT skills, the post-tests scores showed that these differences were found mainly for Romanian and Lithuanian students ($p<.050$); the gain was lower in the former, and the highest in the latter.

When the analysis of gains was repeated using only the data from the three most represented countries (Portugal, Greece, and Romania), because the low number of respondents from Lithuania and Germany could miner the analysis strength, the post-test scores significantly differ from those of pre-tests in both integrated skills ($F = 3.312$; $df = 2$; $p = .038$) and dispositions ($F = 3.224$; $df = 2$; $p = .042$) (Table 19). The gain in CT skills and dispositions was the highest in the Portuguese students, whereas the Romanian students showed the lowest gain (Table 19). The post-hoc tests showed that significant differences existed between students from Portugal and Romania, but not between students from Portugal and Greece, or between Romanian and Greek students.

Co-funded by the Erasmus+ Programme of the European Union

Table 19. Gains in the integrated scores for CT Skills and Dispositions in the experimental group of students enrolled i*n* CTBACs for the three countries with a larger number of participants.

		N	Mean*	Std. Deviation	Minimum	Maximum	F (df = 2)	p
Changes in SKILLS	**Portugal**	*100*	*7.19*	*9.97*	*-1.14*	*1.75*	*3.312*	***.038***
	Romania	*81*	*3.43*	*10.49*	*-1.85*	*1.79*		
	Greece	*63*	*6.13*	*8.99*	*-.91*	*1.88*		
	Total	*244*	*5.67*	*10.20*	*-1.85*	*1.88*		
Changes in DISPOSITIONS	**Portugal**	*100*	*.92*	*3.50*	*-1.61*	*1.72*	*3.224*	***.042***
	Romania	*81*	*-.25*	*3.88*	*-1.82*	*1.46*		
	Greece	*63*	*-.30*	*3.36*	*-1.42*	*1.11*		
	Total	*244*	*.21*	*3.63*	*-1.82*	*1.72*		

*the mean represents the average difference for the post-test and pre-test scores in each country. The higher and positive the mean is, the higher the change in the score recorded after the intervention.

Changes in CT skills and dispositions *per* discipline or country

Germany

The German sample was composed of 22 students enrolled in the Business Informatics Discipline, with an average age of 24.05 years (sd = 5.28; range = 19 - 38). 81.8% of participants were male and 18.2% were female. Students participated in this study within the scope of four Courses: Design patterns (n=8); Economic Aspects of Industrial Digitalization (n=8); Innovation Management (n=4) and Scientific seminar (n=2). Table 20 presents the descriptive results for each scale and subscale.

Table 20. Descriptive statistics for the German population for CT skills and dispositions at the onset of the pilot courses

	Pre-test scores	Mean	Std. Deviation	Minimum	Maximum
SKILLS	Interpretation	11.71	2.68	7.08	16.58
	Evaluation	8.13	1.83	3.30	11.10
	Analysis	10.80	3.26	4.20	16.20
	Inference	12.23	2.66	6.67	17.47
	Explanation	11.34	2.70	6.17	16.50
	Self-Regulation	8.32	1.88	3.86	11.86
	CT skills' integrated score	62.54	13.41	32.86	85.14
DISPOSITIONS	Reflection	5.64	.66	4.67	6.67
	Attentiveness	4.10	1.38	2.00	6.25
	Open-Mindedness	5.13	1.18	2.50	7.00
	Organization	4.82	1.28	1.67	6.67
	Perseverance	5.45	.96	3.67	7.00
	Intrinsic Goal Motivation	5.65	.89	3.50	7.00
	CT dispositions' integrated score	30.78	3.88	25.25	38.92

Co-funded by the Erasmus+ Programme of the European Union

The verification of prior equality between groups according to age or course was not performed because of the unbalance between the groups and due to the small representativeness of the groups for the variable course. Age did not affect the prior equality neither for CT skills nor CT dispositions. The overall pre-test/post-test comparisons presented a positive effect of the intervention in two skills: "*Interpretation*" and "*Analysis*" (Table 21).

Table 21. Difference of means and results of comparison of means in the German experimental group (paired t-test) (n=22)

		Mean	Std. Deviation	Mean difference	t	p
SKILLS	POST_ Interpretation	12.73	2.34	1.022	2,089	***. 049***
	PRE_ Interpretation	11.71	2.68			
	POST_ Evaluation	8.44	1.50	.309	.978	.339
	PRE_ Evaluation	8.13	1.83			
	POST_ Analysis	12.16	2.96	1.359	2.286	***.033***
	PRE_ Analysis	10.80	3.26			
	POST_ Inference	12.05	2.93	-.182	-.305	.763
	PRE _Inference	12.23	2.66			
	POST_ Explanation	12.29	2.98	.947	1.575	.130
	PRE_ Explanation	11.34	2.70			
	POST_ Self-Regulation	8.47	1.88	.149	.349	.731
	PRE_ Self-Regulation	8.32	1.87			
	POST_ INTEGRATED SCORE	66.15	13.29	3.605	1.466	.157
	PRE_ INTEGRATED SCORE	62.54	13.41			
DISPOSITIONS	POST_ Reflection	5.53	.94	-.106	-.480	.636
	PRE_Reflection	5.64	.66			
	POST_ Attentiveness	4.33	1.30	.227	.814	.425
	PRE_ Attentiveness	4.10	1.38			
	POST_Open-Mindedness	4.91	1.27	-.215	-.631	.525
	PRE_ Open-Mindedness	5.13	1.18			
	POST_ Organization	4.95	1.42	.136	.483	.634
	PRE_ Organization	4.82	1.28			

Co-funded by the Erasmus+ Programme of the European Union

		Mean	Std. Deviation	Mean difference	t	p
DISPOSITIONS	POST_ Perseverance	5.18	1.11	-.273	-1.393	.178
	PRE_ Perseverance	5.45	.96			
	POST_ Intrinsic Goal Motivation	5.47	1.24	-.182	.774	.448
	PRE_ Intrinsic Goal Motivation	5.65	.89			
	POST_ INTEGRATED SCORE	30.37	4.68	-.413	-.479	.637
	PRE_ INTEGRATED SCORE	30.78	3.88			

The pedagogical interventions did not influence the dispositions outcome, possibly because of the small number of the sample.

Greece

The Greek experimental sample was composed of 63 students enrolled in the Teacher Education Discipline; this population presented an average age of 24.48 years (sd = 8.42; range = 19 - 54). 63.5% of participants were between 20 and 21 years old. 85.7% of participants are female and 14.3% are male. Students participated in this study within the scope of three courses: Teaching Science Education (54%), Teaching Biological Concepts (34.9%) and Teaching of the Study of the Environment (11.1%). Table 22 shows the descriptive results in each of the subscales and scales for the experimental group at the beginning of CTBACs.

Table 22. Statistics (means, SD, minimum and maximum scores) for CT skills and dispositions in the Greek experimental group before the CTBACs' implementation.

	Dimensions	Mean	Std. Deviation	Minimum	Maximum
Skills	***Interpretation***	12.41	2.61	6.67	18.00
	Evaluation	7.96	1.78	3.20	11.10
	Analysis	11.97	2.51	5.20	17.50
	Inference	13.05	2.37	6.53	17.20
	Explanation	11.38	2.51	5.33	16.17

Co-funded by the Erasmus+ Programme of the European Union

	Dimensions	Mean	Std. Deviation	Minimum	Maximum
Skills	***Self-Regulation***	*8.32*	*2.01*	*3.57*	*12.00*
	INTEGRATED SCORE	*65.09*	*11.54*	*33.48*	*90.51*
Dispositions	***Reflection***	*5.88*	*.88*	*3.67*	*7.00*
	Attentiveness	*4.00*	*1.58*	*1.00*	*7.00*
	Open-Mindedness	*5.06*	*1.20*	*2.00*	*7.00*
	Organization	*4.90*	*1.45*	*1.00*	*7.00*
	Perseverance	*5.49*	*1.11*	*2.67*	*7.00*
	Intrinsic Goal Motivation	*5.24*	*1.20*	*1.75*	*7.00*
	INTEGRATED SCORE	*30.56*	*4.22*	*20.83*	*38.75*

The existence of differences in means in the pre-intervention phase between sexes was not evaluated due to the enormous imbalance between the sizes of both groups. Comparison between Courses was performed only between the two groups with higher number of respondents: Teaching Science Education (n = 34) and Teaching Biological Concepts (n = 22) (Table 23). The comparison reveals significant differences in the baseline scores of students from the two courses in two skills: *Interpretation* and *Analysis.* The differences were the most evident in the CT dispositions scale. In all cases, the averages of students, who attended the course "Teaching Science Education", are higher than the students who attended the course "Teaching Biological Concepts". Finally, one positive and significant relationship was observed between age and the disposition *Intrinsic Goal Motivation* (*r*=.337; p=.007).

Table 23. CT skills and dispositions means at the pre-test in Greek students at the courses of Teaching Science Education (n=34) and Teaching Biological Concepts (n=22).

	PRE-test	Courses	Mean	Std. Deviation	P
Skills	**Interpretation**	**Teaching Science Education**	*13.38*	*2.08*	*.007*
		Teaching Biological Concepts	*11.38*	*2.81*	
	Evaluation	**Teaching Science Education**	*8.13*	*1.60*	*.339*
		Teaching Biological Concepts	*7.60*	*2.17*	

Co-funded by the Erasmus+ Programme of the European Union

	PRE-test	Courses	Mean	Std. Deviation	P
Skills	Analysis	Teaching Science Education	12.51	2.07	.039
		Teaching Biological Concepts	11.07	3.02	
	Inference	Teaching Science Education	13.30	2.20	.218
		Teaching Biological Concepts	12.46	2.82	
	Explanation	Teaching Science Education	11.57	2.20	.498
		Teaching Biological Concepts	11.11	2.72	
	Self-Regulation	Teaching Science Education	8.62	1.99	.067
		Teaching Biological Concepts	7.58	2.03	
	Integrated Score	Teaching Science Education	67.50	9.88	.053
		Teaching Biological Concepts	61.20	13.91	
Dispositions	Reflection	Teaching Science Education	6.02	.87	.086
		Teaching Biological Concepts	5.61	.86	
	Attentiveness	Teaching Science Education	4.38	1.53	.024
		Teaching Biological Concepts	3.43	1.46	
	Open-Mindedness	Teaching Science Education	4.98	1.21	.946
		Teaching Biological Concepts	5.00	1.18	
	Organization	Teaching Science Education	5.31	1.32	.022
		Teaching Biological Concepts	4.41	1.50	
	Perseverance	Teaching Science Education	5.67	1.05	.086
		Teaching Biological Concepts	5.15	1.11	
	Intrinsic Goal Motivation	Teaching Science Education	5.65	1.09	.004
		Teaching Biological Concepts	4.68	1.20	
	Integrated Score	Teaching Science Education	32.01	4.05	.001
		Teaching Biological Concepts	28.28	3.48	

The general pre-test/post-test comparisons (without considering the prior differences) show a positive effect of the intervention on all skills, with students presenting higher means after the intervention. In the case of CT dispositions, there is only a significant difference in *Open-Mindedness* but in an opposite direction than expected, with a higher mean in the moment before the intervention than in the later moment (Table 24).

In general, students enrolled in the course "Teaching Science Education" presented higher gains in CT skills and integrated score following the interventions than the Teaching Biological

Co-funded by the Erasmus+ Programme of the European Union

Concepts' students, except for *Interpretation* (12.49 vs. 13.42). In addition, for dispositions, the mean differences in dispositions' scores were higher for students enrolled at the course "Teaching Science Education" in comparison to students enrolled at the course "Teaching Biological Concepts", with exception of the dimension *Attentiveness* (3.44 vs. 3.96). Nonetheless, these differences recorded in the two experimental groups failed to reach significance. However, because of the prior existing differences between the groups in the covariate, namely for the skills "*Interpretation*" and "*Analysis*" and for three of the six dispositions and the integrated CT dispositions score (as presented earlier) the results must be interpreted with caution.

The effects of the pedagogical interventions were also compared between the experimental and control groups. The experimental condition (control vs experimental) was used as the independent variable, post-measurement as the dependent variable and pre-measurement as a covariate. The confirmation of non-existence of prior differences between the groups in the skills and dispositions dimensions showed that the assumption was met for all but the *Open-Mindedness* dimension, in which the experimental group had a higher previous average than the control group (5.06 vs. 4.34). The experimental group showed higher positive changes in the integrated CT skills score (p=.007), as well as in the dimensions "*Evaluation*" (p=.040), "*Inference*" (p=.042), "*Explanation*" (p=.008), but not in "*Interpretation*", "*Analysis*" or "*Self-Regulation*". A similar pattern was found for the gain in CT dispositions. The differences were higher in the experimental group for the dispositions' integrated score (p=.007) and the domain "*Attentiveness*" (p=.019), but no differences were found between the experimental and control group for the other dispositions.

Table 24. Difference of means and results of comparison of means (paired t-test) in the Greek experimental group (n=63)

Co-funded by the Erasmus+ Programme of the European Union

		Mean	Std. Deviation	Mean difference	t	p
SKILLS	POST_ Interpretation	13.26	2.54	.852	2.469	**.016**
	PRE_ Interpretation	12.41	2.61			
	POST_ Evaluation	8.86	1.99	.903	4.642	**≤.0001**
	PRE_ Evaluation	7.96	1.78			
	POST_ Analysis	13.09	2.91	1.123	4.642	**≤.0001**
	PRE_ Analysis	11.97	2.51			
	POST_ Inference	13.90	2.77	.853	4.642	**≤.0001**
	PRE _Inference	13.05	2.37			
	POST_ Explanation	12.99	2.73	1.616	5.454	**≤.0001**
	PRE_ Explanation	11.38	2.51			
	POST_ Self-Regulation	9.10	1.909	.782	3.890	**≤.0001**
	PRE_ Self-Regulation	8.32	2.01			
	POST_ INTEGRATED SCORE	71.22	13.23	6.130	5.412	**≤.0001**
	PRE_INTEGRATED SCORE	65.09	11.54			
DISPOSITIONS	POST_ Reflection	5.90	.90	.026	-.217	.829
	PRE_Reflection	5.88	.88			
	POST_ Attentiveness	3.84	1.69	-.159	.941	.350
	PRE_ Attentiveness	4.00	1.58			
	POST_Open-Mindedness	4.56	1.47	-.504	2.758	**.008**
	PRE_ Open-Mindedness	5.06	1.20			
	POST_ Organization	5.11	1.22	.206	-1.326	.190
	PRE_ Organization	4.90	1.45			
	POST_ Perseverance	5.56	1.05	.074	-.619	.538
	PRE_ Perseverance	5.49	1.11			
	POST_ Intrinsic Goal Motivation	5.30	1.20	.056	-.433	.667
	PRE_ Intrinsic Goal Motivation	5.24	1.20			
	POST_ INTEGRATED SCORE	30.56	4.22	.300	.709	.481
	PRE_INTEGRATED SCORE	30.26	4.49			

Co-funded by the Erasmus+ Programme of the European Union

Lithuania

The Lithuanian sample is composed of 20 students enrolled in the course *English as Foreign Language*, from the International Relations and Political Science (BA) with an average age of 18.75 years (sd = .639; range = 18 - 20). 65% of the participants were female and 35% were male. Students participated in this study within the scope of the pilot course *English as Foreign Language*. Table 25 presents the descriptive results for each dimension and subdimensions of the instrument used to collect data.

Table 25. Statistics (means, SD, minimum and maximum scores) for CT skills and dispositions in the Lithuanian experimental group before CTBACs' implementation

	Dimensions	Mean	Std. Deviation	Minimum	Maximum
Skills	Interpretation	11.47	2.20	6.83	15.50
	Evaluation	7.98	1.91	3.90	11.50
	Analysis	9.80	3.43	4.50	17.25
	Inference	10.89	3.25	4.53	17.40
	Explanation	10.95	3.09	6.50	16.67
	Self-Regulation	7.32	2.28	3.00	11.19
	INTEGRATED SCORE	58.40	13.83	38.13	88.01
Dispositions	Reflection	5.57	.77	4.00	7.00
	Attentiveness	4.08	1.41	1.50	6.75
	Open-Mindedness	5.51	.80	4.25	6.75
	Organization	4.22	1.66	1.00	6.67
	Perseverance	5.00	1.14	3.00	7.00
	Intrinsic Goal Motivation	5.14	1.16	2.50	7.00
	INTEGRATED SCORE	29.51	4.47	20.58	38.75

The verification of prior equality according to age-groups was not performed because of the unbalance between the two groups. A significant but negative association was found between age and the *Open-Mindedness* disposition (r= -.511; p = .021).

Co-funded by the Erasmus+ Programme of the European Union

The pre-test/post-test comparisons evidenced a positive effect of the intervention in all but one CT skill: "*Evaluation*" (Table 26). In the case of Lithuania, only one course was considered and therefore no variation in the learning strategies were considered.

Table 26. Difference of means and results of comparison of score' means (paired t-test) in the Lithuanian students (n=20)

		Mean	Std. Deviation	Mean difference	t	p
SKILLS	POST_ Interpretation	13.43	2.14	1.958	-3.738	**.001**
	PRE_ Interpretation	11.47	2.20			
	POST_ Evaluation	8.63	1.94	.655	-1.489	.153
	PRE_ Evaluation	7.98	1.91			
	POST_ Analysis	12.63	2.80	2.835	-4.281	**≤.0001**
	PRE_ Analysis	9.80	3.43			
	POST_ Inference	12.60	2.74	1.717	-3.321	**.004**
	PRE _Inference	10.89	3.25			
	POST_ Explanation	13.08	2.50	2.125	-4.959	**≤.0001**
	PRE_ Explanation	10.95	3.09			
	POST_ Self-Regulation	8.58	2.07	1.257	-3.110	**.006**
	PRE_ Self-Regulation	7.32	2.28			
	POST_ INTEGRATED SCORE	68.95	12.63	10.547	4.644	**≤.0001**
	PRE_INTEGRATED SCORE	58.40	13.83			
DISPOSITIONS	POST_ Reflection	5.77	.88	.200	1.092	.289
	PRE_Reflection	5.57	.77			
	POST_ Attentiveness	4.31	1.61	.238	.937	.361
	PRE_ Attentiveness	4.08	1.41			
	POST_Open-Mindedness	5.64	.58	.125	1.022	.320
	PRE_ Open-Mindedness	5.51	.80			
	POST_ Organization	4.48	1.73	.267	1.823	.084
	PRE_ Organization	4.22	1.66			
	POST_ Perseverance	4.67	1.59	-.333	-1.541	.140
	PRE_ Perseverance	5.00	1.14			
	POST_ Intrinsic Goal Motivation	5.08	1.44	-.063	-.260	.798
	PRE_ Intrinsic Goal Motivation	5.14	1.16			

Co-funded by the Erasmus+ Programme of the European Union

		Mean	Std. Deviation	Mean difference	t	p
DISPOSITIONS	POST_ INTEGRATED SCORE	*29.94*	*5.73*	*.433*	*.712*	*.485*
	PRE_INTEGRATED SCORE	*29.51*	*4.47*			

Portugal

The Portuguese experimental group was composed of 100 students enrolled in the discipline *Veterinary Medicine (Integrated Master)*, with an average age of 22.32 years (sd = 4.62; range = 19 - 52). 79% of the participants were female and 21% were male. Students participated in this study within the scope of three pilot courses: *Imaging* (36%), *Deontology* (42%) and *Gynecology, Andrology & Obstetrics* (22%). Table 27 presents the descriptive results for each dimension and subdimension of the CT skills and disposition scales.

Table 27. Statistics (means, SD, minimum and maximum scores) for CT skills and dispositions in the Portuguese experimental group before CTBACs' implementation

	Dimensions	Mean	Std. Deviation	Minimum	Maximum
Skills	Interpretation	*12.76*	*1.87*	*7,50*	*16,75*
	Evaluation	*7.62*	*1.73*	*2,60*	*11,60*
	Analysis	*11.58*	*2.14*	*5,75*	*15,75*
	Inference	*11.72*	*2.34*	*4,93*	*16,40*
	Explanation	*11.04*	*2.42*	*5,17*	*16,17*
	Self-Regulation	*8.56*	*1.71*	*5,14*	*11,71*
	INTEGRATED SCORE	*63.28*	*10.06*	*37,91*	*86,55*
Dispositions	Reflection	*5.57*	*1.01*	*1,67*	*7,00*
	Attentiveness	*3.87*	*1.40*	*1,00*	*7,00*
	Open-Mindedness	*5.91*	*1.08*	*1,00*	*7,00*
	Organization	*4.95*	*1.49*	*1,00*	*7,00*
	Perseverance	*5.31*	*1.18*	*1,67*	*7,00*
	Intrinsic Goal Motivation	*5.32*	*1.12*	*1,50*	*7,00*
	INTEGRATED SCORE	*30.93*	*4.60*	*19,50*	*41,50*

Co-funded by the Erasmus+ Programme of the European Union

The comparison of means between sexes at the beginning of the semester reveals that there are no significant differences between male and female students on the CT skills scales except for the skill "*Analysis*" (p = .035). In all cases, men presented higher averages than women. Regarding the CT dispositions, significant differences were found in "*Organization*" (p = .001), with women presenting a higher average than men. Age did not present a significant relationship with any of the CT skills or disposition dimensions.

On the CT skills scale, comparison between courses showed significant differences in the score means prior to the interventions, specifically in the skills "*Analysis*" (p = .034) and "*Explanation*" (p = .042). In all cases, Gynecology students had higher averages than Imaging students, who in turn had higher averages prior to the intervention than Deontology students. Regarding the *Analysis* and *Explanation* skills, the Bonferroni tests evidenced significant differences only between the Gynecology and Deontology Courses (p = .040 and p = .041, respectively), with Gynecology students presenting a higher average (M = 12.61 and M = 11.97, respectively) than Deontology students (M = 11.23 and M = 10.41, respectively). There are no differences between these two groups and the Imaging students, which have averages of 11.35 and 11.20, respectively. In contrast, there are no significant differences prior to the intervention in CT dispositions that may be associated with the course.

The overall changes in response to the pedagogical interventions (without considering the prior differences detected in some skills depending on sex and course) show a positive effect of the intervention with significantly higher means on all skills' dimensions. Nonetheless, in the case of the CT dispositions, the gains from the intervention are not so clear because, despite the significant overall gain recorded (p = .010) only in 3 of the 6 dispositions the students showed significantly higher scores, and in two of them the results worsened after the intervention even though the differences did not reach statistical significance (Table 28).

Table 28. Difference of means and results of comparison of score' means (paired t-test) in the Portuguese experimental group (n=100)

Co-funded by the Erasmus+ Programme of the European Union

		Mean	Std. Deviation	Mean difference	t	p
SKILLS	POST_ Interpretation	13.85	1.90	1.092	5.028	≤.0001
	PRE_ Interpretation	12.76	1.87			
	POST_ Evaluation	8.70	1.54	1.076	5.983	≤.0001
	PRE_ Evaluation	7.62	1.73			
	POST_ Analysis	12.98	2.14	1.405	6.121	≤.0001
	PRE_ Analysis	11.58	2.14			
	POST_ Inference	13.24	2.03	1.521	6.648	≤.0001
	PRE _Inference	11.72	2.34			
	POST_ Explanation	12.46	2.22	1.420	5.919	≤.0001
	PRE_ Explanation	11.04	2.42			
	POST_ Self-Regulation	9.23	1.50	.677	4.274	≤.0001
	PRE_ Self-Regulation	8.56	1.71			
	POST_ INTEGRATED SCORE	70.47	9.98	7.190	7.211	≤.0001
	PRE_INTEGRATED SCORE	63.28	10.06			
DISPOSITIONS	POST_ Reflection	5.90	.81690	.330	3.590	.001
	PRE_Reflection	5.57	1.01			
	POST_ Attentiveness	3.76	1.43	-.113	-.966	.336
	PRE_ Attentiveness	3.87	1.40			
	POST_Open-Mindedness	5.913	1.30	.005	.042	.966
	PRE_ Open-Mindedness	5.907	1.08			
	POST_ Organization	5.09	1.38	.137	1.215	.227
	PRE_ Organization	4.95	1.49			
	POST_ Perseverance	5.55	1.18	.247	2.899	.005
	PRE_ Perseverance	5.31	1.18			
	POST_ Intrinsic Goal Motivation	5.64	.95	.320	3.655	≤.0001
	PRE_ Intrinsic Goal Motivation	5.32	1.12			
	POST_ INTEGRATED SCORE	31.85	4.54	.916	2.618	.010
	PRE_INTEGRATED SCORE	30.93	4.60			

The differential effect of the interventions performed in the three courses was analyzed, but no differential effects were found for the course over the scores obtained by the students (Table 29).

Co-funded by the Erasmus+ Programme of the European Union

Table 29: Course-related effects in the average gain (±sd) in CT skills and dispositions in the Portuguese experimental group before the CTBACs' implementation

	Dimensions	Imaging (n=36)	Deontology (n=42)	Gynecology (n=22)	F (df = 2)	p	Eta
Skills	Interpretation	*13,37±2.34*	*14.25 ±1.67*	*13.88±1.31*	*2.079*	*0.131*	*0.042*
	Evaluation	*12,6±2.53*	*13.19±1.93*	*13.16±1.81*	*1.224*	*0.299*	*0.025*
	Analysis	*8.41±1.54*	*8.90±1.47*	*8.78±1.38*	*1.952*	*0.148*	*0.039*
	Inference	*12.84±2.28*	*13.35 ±1.94*	*13.69±1.75*	*1.356*	*0.263*	*0.027*
	Explanation	*12.05±2.37*	*12.43±2.11*	*13.17±2.07*	*1.669*	*0.194*	*0.034*
	Self-Regulation	*9.25±1.71*	*9.18±1.37*	*9.32±1.46*	*0.013*	*0.987*	*0.000*
	INTEGRATED SCORE	*68.55±11.65*	*71.30 ±9.08*	*72.00±8.52*	*1.568*	*0.214*	*0.032*
Dispositions	Reflection	*5.94±0.90*	*5.87±0.81*	*5.92±0.72*	*0.279*	*0.757*	*0.006*
	Attentiveness	*4.13±1.36*	*3.63±1.43*	*3.41±1.45*	*1.680*	*0.192*	*0.034*
	Open-Mindedness	*6.21±0.94*	*5.64±1.44*	*5.92±1.44*	*1.821*	*0.167*	*0.037*
	Organization	*5.31±1.35*	*4.91±1.34*	*5.06±1.51*	*0.278*	*0.758*	*0.006*
	Perseverance	*5.78±1.23*	*5.33±1.22*	*5.62±0.93*	*1.174*	*0.313*	*0.024*
	Intrinsic Goal Motivation	*5.80±1.01*	*5.48±0.90*	*5.68±0.96*	*0.613*	*0.544*	*0.013*
	INTEGRATED SCORE	*33.15±4.74*	*30.85±4.40*	*31.62±4.12*	*0.614*	*0.543*	*0.013*

Romania

The Romanian experimental group was composed of 81 students from the *Business and Economics* discipline, with a mean age of 26.58 years (sd = 10.04; rank = 19-52). 65% of the participants are between 19 and 23 years old. 84% of the participants are female and 16% male. Students participated in this study within the scope of three Courses: *Business Communication* (38.3%), *Pedagogy and Didactics of Financial Accounting* (39.5%) and *Virtual Learning Environments in Economics* (22.2%). Table 30 shows the descriptive results in each of the dimensions of CT skills and dispositions.

Table 30. Statistics (means, sd, minimum and maximum scores) for CT skills and dispositions in the Romanian experimental group before CTBACs' implementation

	Dimensions	Mean	Std. Deviation	Minimum	Maximum
Skills	Interpretation	13.28	2.29	8.25	18.00
	Evaluation	8.92	1.81	3.90	11.80
	Analysis	13.26	2.68	6.75	17.40
	Inference	13.40	2.59	6.87	17.60
	Explanation	12.20	2.84	6.00	18.00
	Self-Regulation	9.15	1.91	5.10	12.00
	INTEGRATED SCORE	70.20	12.26	42.41	92.65
Dispositions	Reflection	5.91	.80	4.00	7.00
	Attentiveness	4.11	1.34	1.75	7.00
	Open-Mindedness	5.18	1.28	2.25	7.00
	Organization	5.10	1.22	2.33	7.00
	Perseverance	5.54	1.08	2.33	7.00
	Intrinsic Goal Motivation	5.59	.91	3.00	7.00
	INTEGRATED SCORE	31.44	4.15	19.50	39.75

The comparison of means between sexes in the pre-intervention phase reveals that there are no significant differences between male and female students neither in the dimensions of the

Co-funded by the Erasmus+ Programme of the European Union

CT skills or dispositions, except for the skill "*Analysis*" (11.76±2.82 vs. 13.55±2.57). respectively) and the disposition "*Perseverance*" (4.97±1.23 vs. 5.65±1.03).

The comparison between the three pilot courses reveals significant differences in the baseline scores of students from different courses in the skill "*Interpretation*" and in the disposition "*Perseverance*". In either case, the Bonferroni post-hoc test confirms that the differences exist between the *Pedagogy and Didactics of Financial Accounting* and the *Virtual Learning Environments in Economics* courses. In addition, in both cases, *Virtual Learning* students present a higher average score in *Interpretation* (14.35 vs 12.67) and *Perseverance* (6.00 vs 5.20) than students in the *Pedagogy and Didactics of Financial Accounting*.

Finally, regarding the age, two positive and significant relationships are observed, in the same skill and disposition mentioned before: *Interpretation* (p=.004) and *Perseverance* (p=.020).

Overall pre-test/post-test comparisons (without considering prior differences) show a positive effect of the intervention on only two skills (*Explanation* and *Interpretation*) and on the integrated CT skills score (Table 31).

The GLM- Univariate ANCOVA test was computed to estimate the gains in the pilot courses [*Business Communication* (n=31), *Pedagogy and Didactics of Financial Accounting* (n=32) and *Virtual Learning Environments in Economics* (n= 18)]. Nonetheless, some care should be taken when interpreting the differences identified, as the assumptions of non-existing prior differences between the groups in the covariate was not fulfilled for the skill, *Interpretation* and disposition *Perseverance* (see previous paragraphs).

In general, the intervention-associated gains in CT skills did not differ between the three pilot courses. In the CT dispositions' integrated score, it was higher (p=.017) for the *Business Communication* and the *Virtual Learning Environments in Economics* students (32.45±4.613 and 32.25±3.78, respectively) than those enrolled in the *Pedagogy and Didactics of Financial Accounting* course (29.38±4.34). Regarding the gains in dispositions, the differences between

Co-funded by the Erasmus+ Programme of the European Union

the courses were observed for: *"Attentiveness"* (p=.028), the higher gains recorded in the *Business Communication* students compared with the *Virtual Learning Environments in Economics* or the *Pedagogy and Didactics of Financial Accounting* (4.45±1.31 vs. 3.73+1.21 vs. 3.89±1.38, respectively); "*Open-Mindedness*" (p=.047) the students in the *Business Communication* and the *Pedagogy and Didactics of Financial Accounting* courses presenting higher gains than students in *Virtual Learning Environments in Economics* (5.14±1.40 and 5.22±1.30 vs. 4.47±1.33 vs); and in "*Intrinsic Goal Motivation*" (p=.009). In the latter, both the students in *Business Communication* and *Pedagogy and Didactics of Financial Accounting* courses showed higher gains (5.92±1.02 and 5.75±.76) compared with the students in *Virtual Learning Environments in Economics* (5.11±.91).

Table 31. Difference of means and results of comparison of score' means (paired t-test) in the Romanian students (n=81)

		Mean	Std. Deviation	Mean difference	t	p
SKILLS	POST_ Interpretation	14.09	2.16	.809	3.346	***.001***
	PRE_ Interpretation	13.28	2.29			
	POST_ Evaluation	9.21	1.68	.295	1.571	.120
	PRE_ Evaluation	8.92	1.81			
	POST_ Analysis	13.55	2.39	.290	1.043	.300
	PRE_ Analysis	13.26	2.68			
	POST_ Inference	13.85	2.49	.456	1.727	.088
	PRE _Inference	13.40	2.59			
	POST_ Explanation	13.43	2.43	1.237	4.092	***≤.0001***
	PRE_ Explanation	12.20	2.84			
	POST_ Self-Regulation	9.49	1.82	.346	1.886	.063
	PRE_ Self-Regulation	9.15	1.91			
	POST_ INTEGRATED SCORE	73.63	11.67	3.431	2.944	***.004***
	PRE_INTEGRATED SCORE	70.20	12.26			
DISPOSITIONS	POST_ Reflection	5.81	.93	-.103	-1.038	.303
	PRE_Reflection	5.91	.80			

Co-funded by the Erasmus+ Programme of the European Union

		Mean	Std. Deviation	Mean difference	t	p
DISPOSITIONS	POST_ Attentiveness	4.04	1.31	-.065	-.474	.637
	PRE_ Attentiveness	4.11	1.34			
	POST_Open-Mindedness	4.89	1.38	-.290	-1.969	.052
	PRE_ Open-Mindedness	5.18	1.28			
	POST_ Organization	5.30	1.15	.202	1.407	.163
	PRE_ Organization	5.10	1.22			
	POST_ Perseverance	5.58	1.08	.037	.293	.771
	PRE_ Perseverance	5.54	1.08			
	POST_ Intrinsic Goal Motivation	5.56	.98	-.031	-.307	.760
	PRE_ Intrinsic Goal Motivation	5.59	.91			
	POST_ INTEGRATED SCORE	31.19	4.53	-.250	-.580	.564
	PRE_INTEGRATED SCORE	31.44	4.15			

3. Discussion

The piloting CTBAC involved 609 students from the five partners' countries, enrolled at different courses within a given programme or discipline (revisit Table 1). This number is considerably higher than the initially proposed in the application form of the Think4Jobs project (150 students were envisaged then to be enrolled in CTBACs courses) and result, at least in part, from the discordance of students within the same Institution and academic year, to learn according to different strategies or approaches. Therefore, in some countries, where the same course was not offered in two consecutive semesters to a different group of students, it was not possible to gather enough students to form a control group. It is possible that students perceived the proposed differences in the learning process either as unfair or detrimental - as leading to different learning levels -, or possibly originating situation that increased their workload without guaranteeing equivalent learning, failing to perceive the associated gains.

Control groups were only obtained in Greece (40 students enrolled in the Teaching Science Education course, from the Teacher Education Programme), and in Portugal, for the curricular

Co-funded by the Erasmus+ Programme of the European Union

traineeships of Veterinary Medicine Programme (n=9), that were developed at *Hospital Veterinário do Atlantico*.

Out of the 609 students enrolled in the piloting CTBAC, only 53,7% (n=286) responded to the pre-test and post-test questionnaires aiming at scoring the students' CT skills and dispositions before and after the pilot interventions. Nonetheless, this number is sufficient to support and validate the different pedagogical interventions implemented per country. In the application we proposed to assess the evolution of the CT skills and dispositions in three different moments of the piloting CTBAC. Even though the questionnaires used to evaluate CT in students were applied at the three proposed moments (pre-test, intermediary, and post-test), the number of students that responded decreased from the pre-test to the post-test measurement, particularly in the courses implemented during the first semester of the academic year 2021-2022. Therefore, to avoid losing information it was decided to use only the responses to the first and final survey moments from the first semester, and to pass the questionnaire twice, in courses developed in the second semester during the academic year 2021-2022.

Even though the request to complete the questionnaires was sent to students via the CTBACs´ Moodle (the platform provides email notification for the tasks and activities) it had a low impact on respondents' response [44]. As no consequences existed for the non-completion of the tasks (responding to questionnaires), nor any money (or voucher or any other benefit for respondents) involved, students did not feel compelled to fill the questionnaires. This problem in engaging students to respond to a variety of feedback questionnaires has been reported earlier [24, 45]. It has been suggested that the weak responses rates to questionnaires in higher education contexts may be related to the fact that students do not anticipate a meaningful retribution of the gathered data in an issue that might be of interest for their personal goals. Therefore, in the absence of a reinforcing stimulus triggering their engagement, which could introduce a conditioning bias [44], the number of respondents tended to remain low and reduced each time the same questionnaire was applied. The

number of questions was reduced to minimize the burden of length, the importance and goals of the questionnaires (validate the instructional strategies effectiveness) was stressed, and the main concepts (CT skills and dispositions and how they are used by professionals) were explained at the onset of the course, along with recommendations regarding the completion of the *Google Forms*, in an attempt to engage students to respond to the questionnaires. We are not aware of the exact reasons for the low response rate to questionnaire, but we believe that respondents were true and willing to provide an unbiased report about their perceived CT skills and dispositions.

The number of respondents was unevenly distributed per country, but the questionnaires were filled voluntarily, without impositions or consequences for non-respondent students. Nonetheless the total number of paired questionnaires was sufficient to run a reliable global statistical analysis and to compare the CT skills and dispositions scores among the three most represented countries (Greece, Portugal, and Romania).

The baseline analysis evidenced the existence of previous differences in the average and scores in the integrated CT disposition scale according to the sex; these differences were restricted to the dimension *Open-mindedness*. Differences were also found for CT skills' integrated score and skills dimensions between countries

Age was positively related with the scores for CT skills, but in the case of the CT dispositions, only for the integrated score and three of the six dimensions reached statistical significance. An age effect over CT skills has been reported before [46]. However, we cannot discard the hypothesis that the age effect might mask other external factors, such as the effects of the country or the discipline. Often the age-effect over CT is less visible when there is a short age interval within the population, and more pronounced when different maturity levels across aging [47] and students in different levels of a programme are compared [48]. The age distribution in students composing the experimental/intervention group was skewed, but despite having students more advanced in age, the predominant students were aged between

Co-funded by the Erasmus+ Programme of the European Union

19 and 21 years (75,6%), with 18,2% of students aged between 22 and 28, and only close to 6% of them aged above 38 years.

In this study, as each country work with a different discipline or programme, the effect of the two overlap. Therefore, the country effects found overlap with possible effects arising from the different fields of the Programmes monitored, as each country represented a different discipline. Following the Project design, this relation cannot be distinguished.

The differences unveiled by the baseline analysis advises that when comparing the results between sex and countries, the existent pre-test/post-test differences should be taken into consideration. The study of the interventions' effects should be assessed separately for each sex and country, to evaluate the true effectiveness of an intervention. However, because of the unbalanced sexes in the population this comparison was not performed for this report. Due to the unbalanced representativeness of the five countries, comparisons between countries were restricted to the three more representative.

CTBACs-associated gains in CT skills and dispositions

The pre- and post-test scores comparison showed that the interventions delivered in the piloting CTBACs allowed students to enhance their integrated CT skills but in contrast the effects on CT dispositions were reduced and often did not reach significance. This supports the idea that dispositions may be harder to change over short time periods (as a semester) as it requests a continuous and focused practice and a continuous exercise of reflection so they can be gradually absorbed. In addition, since dispositions are attitudinal traits, they request an intrinsic willingness and effort to cultivate them, while CT skills represent procedural behaviour that may be incorporated in response to a trigger (training). Besides, different dispositions may be more easily nurtured than others, which could explain why the interventions originated gains in some dispositions but not in others, which may also reflect differences among Programmes/Disciplines.

Co-funded by the Erasmus+ Programme of the European Union

To better assess the impact of CTBACs on CT development, the skills and dispositions gains were compared between the three main representative countries and within each country, as also the strategies implemented, and the CT skills and dispositions targeted with the pedagogical interventions differ among countries/programmes.

The average gain in the CT skills and dispositions integrated scores was higher in Portuguese students, followed by the Greek students. This observation may result from the fact that in general, Romanian students departed the CTBACs scoring higher, or it could also result from differences in the instructional approaches used in the interventions (for details refer to the description of Portuguese interventions for the courses in [1]). For instance, the activities implemented in Veterinary Medicine had a tighter framework that may contribute to the slightly higher gains.

Germany

The University of Emden/Leer (HSEL) and Orgadata AG (Orgadata) implemented two Critical Thinking Apprenticeships Curricula (CTAC) in the academic year 2021-2022 in the discipline of Business Informatics. In particular, two courses were implemented in the winter semester of 2021-2022, and the other two courses “Innovation Management” and “Scientific seminar” in the summer semester of 2022. The “Design patterns” course was a compulsory course offered to apprentices at Orgadata, while the other courses were elective courses offered to students at HSEL. The first two courses were conducted as described in Mäkiö and colleagues [1] and taught in class over a period of 16 weeks, 1.5 hours per week. The other two courses were implemented as described in Mäkiö & Mäkiö [49]. The students who attended these courses formed the experimental group of the planned interventions. As only a small number of students participated in the courses and the self-assessment surveys, the sample size was too small for statistical conclusions.

Due to their professional experience, both in business and education, HSEL and Orgadata educators were aware about the necessity to teach CT skills and dispositions to their students.

Co-funded by the Erasmus+ Programme of the European Union

Thus, they addressed CT skills such as *Interpretation*, *Analysis*, *Evaluation* and *Explanation* in these courses. Pre-test/post-test comparisons point to an overall positive impact of the interventions on the development of the skills *Interpretation* and *Analysis,* but not in *Evaluation*.

CT dispositions such as *Open-mindedness*, *Analyticity*, *Systematicity*, and *Self-confidence* were also addressed in all the courses. Nonetheless, the pre-test/post-test surveys failed to reveal significant changes in students CT dispositions. In addition, a decrease in the mean scores of most subscales was noted. At first view, this was surprising, as some studies among university students in China have shown significant improvements in CT dispositions (e.g., [50]). The positive effects reported in that study [50] may be attributable to the specificity of the subject or relate with differences in the instrument used to assess the CT dispositions. More surprising is the decrease (small) in the mean scores of most subscales, even if devoid of significance. This may be the case especially when the students had relatively high initial levels of dispositions (> 3.5 in 7-point Likert scale). Moreover, changing ingrained thinking habits and patterns of thought requires a lot of time and repeated practice see [51]). Accordingly to Halpern [51] "*it seems clear that the ability to think clearly and the disposition to engage in the effortful process of thinking are the most critical components of* [students] *education. The enhancement of critical thinking skills is also the most challenging and personally rewarding task in which psychologists and educators can engage*" (p.455).

Moreover, dispositions are dependent of a motivational factor that determines whether attitudes are demonstrated [48]. The results obtained are consistent with some studies reporting little and even non-significant changes in CT dispositions, which has been associated with differences in the surveys' sensitivity to detect small variations in attitudes, or with the fact that interventions could have provided small increases in CT dispositions, but these were not sufficiently large to present statistical significance [48]. Contrastingly, other studies showed a clear positive impact of interventions targeting the development of CT dispositions in Higher Education students [52]. However, one must be aware of the difficulties in

Co-funded by the Erasmus+ Programme of the European Union

establishing comparisons among studies reporting the usefulness of pedagogical interventions to foster CT dispositions in students, as often the instruments used to assess are very different between studies and tackle different sets of CT dispositions. In the present study, a CTBAC semester seems not to be sufficient to reach clear, significant positive change in CT dispositions.

Greece

The University of Western Macedonia (UOWM) in collaboration with the Elementary Experimental School of Florina (EESF) implemented three Critical Thinking Apprenticeships Curricula (CTAC) during the academic year 2021-2022 in the discipline of Teacher Education. In particular, during the winter semester of 2021-2022, three courses were implemented and the students attending the courses constituted the experimental group of the designed intervention described in IO4. The three courses were “Teaching Biological Concepts” (n=83), “Teaching Science Education” (n=62) and “Teaching of the Study of the Environment” (n=12). All three courses are mandatory optional/elective courses offered by the Department of Primary Education at UOWM, apart from the course “Teaching Biological Concepts”, which is offered to students by the Department of Early Childhood Education at UOWM. During the spring semester 2021-2022, the course “Teaching Science Education” was implemented with new students (n=85), which composed the control group of the current study. All courses were implemented with the duration of 13 weeks. The course sessions were implemented once per week with a duration of three hours each. Thus, for each course, one three-hour intervention per week was implemented. Still, as described in the design of the CTAC in Mäkiö and colleagues [1] for the two courses, namely “Teaching Science Education” and “Teaching of the Study of the Environment”, the first part of the course was theoretical (around 5 weeks) and then the student-teachers designed and implemented their teaching learning sequences in primary education schools. On the contrary, student-teachers attending the “Teaching Biological Concepts” designed their teaching learning sequences; nevertheless, these were never implemented in school contexts.

Co-funded by the Erasmus+ Programme of the European Union

As said, the three courses were implemented as described in Mäkiö and colleagues [1]. In the "Teaching of the Study of the Environment" the case studies were implemented in class instead of the Moodle as originally proposed. This discrepancy to the initial CTAC design is due to the small number of students that chose the course, and which allowed the instructor to devote more time for in class discussion and applications. During the implementation of the CTAC, it proved to be quite challenging for the instructors to engage the total number of student- teachers in the pre-post measurements of the CT skills and dispositions. In particular, from the total 157 students attending all three courses, only 63 completed both measurements. No other problem regarding the implementation of the courses was reported by the three instructors.

Both UOWM and EESF recognized the importance of certain skills and dispositions that student-teachers should improve during practicum. Thus, Critical Thinking Blended Apprenticeships Curricula focused more on specific skills such as *Analysis*, *Inference*, *Evaluation* and *Self-regulation*. Pre-test/post-test results suggest an overall positive effect of the intervention on all skills, including skills that were not explicitly targeted during the preparation of curricula and activities (i.e., *Interpretation* and *Explanation*). Though not targeted explicitly, the ability to identify the significant and insignificant factors of a given situation (i.e., *Interpretation*) as well as to justify reasoning (i.e., *Explanation*) were also part of the activities that teachers-students were engaged during the semester. Further, after the implementation of the Critical Thinking Blended Apprenticeships Curricula, course instructors reported a relatively good level of participation and perceived improvement in students' CT skills and dispositions. Though instructors perceived changes in CT skills during the semester, some of them expected to see statistically significant changes in some skills (e.g., *Reflection* for the Teaching Science Education class) due to instructor's focus on perceived importance of specific skills and dispositions. Following a different approach on the measurement of reflection, described elsewhere [53], we identified that student-teachers' attending the course "Teaching Science Education" improved their reflective thinking. Still, such a finding was not revealed in the current study, where the shortened CTSAS questionnaire was

Co-funded by the Erasmus+ Programme of the European Union

employed. Overall, our results allow us to argue that during the semester student- teachers improved their CT skills.

As far as the dispositions is concerned, activities were designed to enhance *Open-mindedness*, *Systematicity*, *Self-confidence*, *Inquisitiveness*, and *Cognitive Maturity*. Pre-test/post-tests revealed no statistically significant change in student- teachers' CT dispositions. Moreover, we noticed a decline in the mean score of the subscale *Open-mindedness*. Although these results were not the anticipated ones, they were not surprising either. First, student- teachers scored relatively high in the pre-measurement (the subscales mean was higher than 3.5 in a 7-point Likert scale) indicating that the above dispositions already existed and could be further nurtured. However, since CT dispositions are considered relatively stable, they might need a systematic and continuous effort towards their further enhancement involving numerous courses during the 4-year undergraduate studies. Further, these results are consistent with similar studies in the literature, reporting only few statistically significant changes in undergraduate students' CT dispositions, including their decrease (see, [53, 54]). Considering that dispositions indicate one's willingness and tendency to think critically, they may be considered as even more important than skills. However, cultivating dispositions seems to be a challenging task, which should supported by a critical thinking culture, which entails teaching in this direction. In detail, classrooms as such should privilege activities that focus both on emotions and cognition in the light of interaction with meticulous and continuous practice on CT [55]. This kind of approach is probably more time-consuming, requiring more time than a semester and should be instilled across the curriculum.

In general, differences in CT skills and dispositions identified between the two courses included in the analysis, namely "Teaching Science Education" and "Teaching Biological Concepts" are attributed to the different design of the courses. Among the most striking differences were the different instructional approaches employed by the instructors as well as the implementation of students' Teaching Learning Sequences in real life contexts for the "Teaching Science Education" (i.e., school settings) course.

Co-funded by the Erasmus+ Programme of the European Union

The effects of the interventions were also compared between the experimental and the control groups. The results revealed that the experimental group had a greater increase in terms of integrated scores regarding both skills and dispositions. This finding is reasonable considering that in all three courses of the experimental group, explicit instruction of CT was implemented at the beginning of the semester. Moreover, previous research has indicated that explicit instruction of CT favours students CT development (e.g., [56]). In addition, Heijltjes and colleagues [57] indicated that combining explicit instruction with practice benefits students more in comparison to other instructional approaches such as implicit instruction of CT. Still, when each skill or disposition was examined, our findings indicated that only some skills were improved in favour of the experimental group, such as "*Evaluation*", "*Inference*" and "*Explanation*". These skills were among the anticipated CT outcomes of the course "Teaching for Science Education". Although, the course focused also on the CT skills of "Analysis" and "Self-regulation", no improvement was identified on these skills. We can argue that specific CT skills were improved in favour of the experimental group because the instructor persisted more during instruction on these skills through the learning strategies and activities students were engaged in. To illustrate, after the implementation of their Teaching Learning Sequences (TLS), teacher students presented their original designs and the TLS they had actually implemented. They reflected on the CT skills and dispositions they employed as well as on the changes they would implement in order for their TLSs to be more successful. As far as the distinct dispositions are concerned, our findings indicate that only *Attentiveness* was improved at the experimental group, namely students' willingness to focus and concentrate; to be aware of surroundings, context, consequences, and potential obstacles. This finding can be justified by the fact that before the implementation of their TLSs, the students in the experimental group were requested to prepare a video displaying that the experiments they would implement in class were realistic and provided the expected outcomes in terms of students' knowledge acquisition. For student- teachers to be successful in this activity, they would have to consider the context, the school surroundings as well as the consequences the experiment could bring on students depending on how successfully it would be implemented.

Co-funded by the Erasmus+ Programme of the European Union

This activity was not implemented in the course for the control group. Hence, we assume that the difference in this disposition was expected. Overall, considering the implemented course for the experimental and the control group, it can be assumed that the different activities implemented as well as the explicit instruction of CT might be the reason for the identified changes between the two groups both in terms of CT skills as well as in terms of CT dispositions.

Due to the number and length of the designed activities we expected to observe a greater development of CT skills and dispositions. However, the design and implementation of activities came with some limitations. First, the winter semester of 2021-2022 was the first semester of face-to-face teaching in Greece since the declaration of the pandemic and thus, we have concerns regarding student-teachers readiness to adhere systematically to classes and to participate in all the designed activities. However, course's instructors reported a relatively good level of teacher-student participation in classes, despite the difficulties (e.g., completing the pre-post measurements, meeting some deadlines, recovering from Covid-19, etc.). Another limitation lie in the type of activities. The majority of the activities took place in classes during teaching and required the active involvement of student- teachers (e.g., critical discussion, brainstorming, work in groups, etc.). These activities might not actively engage some –introverted– students who face difficulties with public speaking or avoid expressing their opinion in class, despite the good and supportive climate reported by instructors. In this light, students may not also be familiar with this kind of interactive instruction where their active participation is –in a way– mandatory and constantly promoted in classroom settings. Therefore, some of them may need more time to adapt and actively engage in the process of active learning.

Lithuania

Vilnius University implemented one CTBAC - the *English for Academic Purposes and Research* - a mandatory English language course for specific purposes delivered by the Institute of

Co-funded by the Erasmus+ Programme of the European Union

Foreign Languages of the Faculty of Philology at Vilnius University. The implementation of the CTBAC was carried out in collaboration with the Public Service Language Centre, under the discipline of *International Relations and Political Science* BA at Vilnius University; it was delivered to the 1st year students of the study programme *International Relations and Political Science* BA at the Institute of International Relations and Political Sciences of Vilnius University. The curriculum was implemented in the two semesters of the academic year 2021-2022 (between September and May), as described by Mäkiö and colleagues [1]. Originally, one more CTBAC was planned under the name of *English for Academic Purposes and Research,* to be implemented within the discipline of *Childhood Pedagogy*. It was successfully implemented with the 1st year students of the Faculty of Philosophy of Vilnius University, in the autumn semester of the academic year 2020/2021. The students were pre-tested at the very beginning of CTBAC but, unfortunately, they failed to take the post-test. Particularly, students were provided with an editable version of the survey questionnaire, which allowed them to change the content and the scales of the items; their post-test answers were therefore considered not valid, and thus they were not statistically processed, having been excluded from the final evaluation and the final report.

The overall doctrine of the curriculum was based on task-based and action-oriented approaches (AoA) following the updated version of the Common European Framework of Reference for Languages: Learning, teaching, assessment [58] and a new vision of foreign language teaching as described by Piccardo and North [59]. This approach has been systematically implemented by the Institute of Foreign Languages of the Faculty of Philology of Vilnius University in Foreign Language Teaching starting with the academic year of 2019/2020 [60]. The action-oriented approach rests on several pillars that are at the core of the CTAC: the learner as a social agent, language activities performed in a particular social context, and real-life, action-oriented tasks [59]. Consequently, the CTBAC curriculum was student-centred, and the tasks designed for the students were action-oriented, intended to be authentic, employing problem-solving and real-life situations.

Within the framework of the curriculum, students were meant to familiarize themselves with the requirements of academic writing. For this purpose, two different tasks were proposed. Firstly, they acquired theoretical knowledge about the valid structure of scientific research. After adequately understanding the genre, they were asked to submit a research proposal to prepare and anticipate the next task. The writing task was also directed at the development of students' CT skills. Students had to analyse research articles, provide insights into possible future research and propose a plan for their own research. Thus, the task targeted the CT skills *Analysis*, *Explanation*, *Interpretation*, *Inference*, *Evaluation*. Secondly, at the end of the first semester, an international conference was simulated, where each student was expected to contribute with an individual presentation. The same activity was organized at the end of the 2nd semester, but this time students worked in teams. The activity was structured as an open event where other students or teachers could participate. Both presentations (i.e., research proposal and the congress presentation) were followed by a Q&A session, where the presenters responded to the audience's questions, comments, or suggestions. The tasks started from a well-defined, authentic scenario based on the general topics mentioned in the course description. In terms of CT, the studies tested the intellectual traits discussed above. The individual and team presentations students had to perform were aimed at the enhancement of CT skills and dispositions. Students had to objectively and cohesively present the data obtained from the research articles they had to analyse. Students were expected to resort to critical analysis of the previous research findings, to identify and define key concepts within a selected theoretical framework, compare research findings across at least a few research studies, provide specific examples, draw conclusions, and raise various implications. The students' ability to compare the research findings from the articles and benchmark them with the social, political, or economic contexts, was also of great relevance. Thus, we can state that an explicit focus on the development of students' CT skills (*Analysis*, *Interpretation*, *Inferencing*, *Comparison*) was maintained during the CTBAC.

The analysis of the gains associated with the implementation of CTBACs evidenced students' CT development. The interventions implemented led to an increase of the integrated CT skills

Co-funded by the Erasmus+ Programme of the European Union

score that reached more than 10 points, as well as a significant improvement in all the CT sub-skills excepting for *Evaluation*. In this sub-skill, albeit the post-test score was higher than the pre-test, the difference did not reach significance.

Nevertheless, the pre-test/post-test comparisons indicated that the CT dispositions were not significantly improved by the piloting interventions. Where the dispositions progress could be observed, it is still inconsequential. This can have different reasons. Firstly, the fact that it is harder to change attitudes than procedures. Attitudes request a deeper understanding and commitment to the CT principles and values, and might be highly impacted by the individual's mood, interests and choices [61]. It can also partly explained by subjective unpredictable conditions. The post-test was conducted towards the end of the academic year, when usually stress and the ability to focus on certain tasks deteriorates. To mitigate the risk, we might consider an alternative strategy. The competitiveness increases in the 2nd semester when students get marks instead of pass/fail assignments (1st semester). Traditionally, marks constitute a stimulus but might put extra pressure on students, with visible consequences on their willingness to enforce and make use of their CT dispositions. The hypothesis was not tested, and we do not have enough data to support it. The collected data imply that CT dispositions slightly deteriorated towards the end of the academic year. The findings appear to suggest that the tested curriculum requires a better redistribution of tasks throughout the two semesters to inspire learners to keep their interest in the topics under analysis and on tasks that they have to perform. Overall, our results allow us to argue that during the semester students improved their CT skills.

Portugal

The piloting courses were implemented as proposed in IO3 [1], albeit with minor differences. In *Imaging*, only two of the three proposed interventions were developed with the students, because of the number of national holidays overlapping with classes (we lost in total the equivalent to 2 weeks of classes). In the *Deontology* and *Gynecology, Andrology and Obstetrics*

Co-funded by the Erasmus+ Programme of the European Union

courses, all the three planned activities per course were developed. In the former, the activities were composed of the analysis of a dilemmatic situation followed by the voicing and discussion of different points of view [1], while in the latter the activities focused on the analysis of a clinical condition, the screening of differential diagnosis and the decision-taking about the best course of action for a particular condition in a pet with particular characteristics [62].

In all the three piloting courses, all students preferred using the pilot CTBACs as a learning method rather than creating a control group. The creation of a control group outside the courses was not possible because in one academic year these courses are offered only once. Since the content and the background in Veterinary Medicine Programmes in other Universities are different, it was decided not to organize a control group with outsider students. Therefore, the decision was made to create a small control group composed of students in curricular apprenticeship at *Hospital Veterinário do Atlântico*.

The engagement of most students to respond to the three questionnaires as planned and requested during the CTBACs implementation was difficult as stated earlier in this document. Therefore, the representativeness of students in the pre-test/post-test pairing are often below 50% of the original students enrolled in the activities.

The results of the average gains showed that the pilot interventions in the CTBACs contributed to changes in CT skills and dispositions. In general, the sex did not affect the scores for CT skills and dispositions except for *Analysis* and *Organization*, the former recording higher average scores in men than women, and the latter showing an inverse pattern. Gains in CT skills were reported in all the three pilot courses, as well as regarding three of the six CT dispositions (*Reflection*, *Perseverance*, and *Intrinsic Goal Motivation*) and the dispositions' integrated score.

Following the general pattern, the average gains in the scores were higher for the integrated CT skills (close to 7.2 points) than for dispositions (around 1 point). The gains in CT skills after

Co-funded by the Erasmus+ Programme of the European Union

the CTBACs closely match the skills proposed as outcomes for the learning strategies described in IO3 [1, 63], suggesting a correct alignment of the interventions with the projected CT outcomes. A similar result was not found for the CT dispositions, sparking our reflection.

Data from the Portuguese students showed that the interventions designed for the Veterinary Medicine Programme triggered gains in some of the CT dispositions praised by the labour market, namely *Reflection, Perseverance* and *Intrinsic Goal Motivation.* While in the SENCTDS scale the construct of the former disposition closely embeds the Facione' conceptualization of CT dispositions and the understanding of a reflective sceptic attitude, the last two dispositions encompass a mix of different attitudes combined in new concepts. According to Quinn et al' conceptualization [3], these dispositions configure positive traits or attitudes necessary both in academic contexts and in the workforce. *Perseverance* represents resilience, motivation to persist through demanding tasks, to perform well in the job and the desire to progress. The *Intrinsic Goal Motivation* represents the ability to be positive and enthusiastic towards an assignment or a problem, or the process of learning and search for solutions; it also includes the internal drive to look for answers independently from any rewards [3].

Perseverance, understood as the inclination to reflect on one's behaviour or motivation, has been associated with superior real-world decision making and the ability to reach sound judgements [3]. On another hand, *Intrinsic Goal Motivation* reflects the student's curiosity, mastery and intrinsic satisfaction in the learning process [3], and thus the self-regulatory attitude we long to develop in HE students [64].

Even though differences in particular CT Skills existed in students before implementing the pilot courses (namely in *Analysis* and *Explanation*), which showed higher scores in students enrolled in courses located at later years in the plan of Veterinary Medicine Programme, the comparison of the gains in CT skills and the respective dimensions, did not differ between the three courses. Also, there were no differences in the average gains in students enrolled in the three courses. This finding was a surprise, as even though a general framework was used, the

Co-funded by the Erasmus+ Programme of the European Union

complexity of the activities, the strategies, the form they were implemented, and the outcomes proposed for the activities differed among the three CTBACs.

Such results suggest that the intervention' intentionality more than the type of strategy may efficiently enhance the students´ CT. The key issue may be to transpose the focus from the product (knowledge) into the process (reasoning) and thus shifting into a constructivist approach to learning.

Romania

During the three courses conducted in Romania (Pedagogy and didactics of financial accounting, Virtual Learning Environments in Economics and Business Communication), trainers from labour market organizations presented various real-life case studies to the students. Thus, on the basis of the theoretical information previously acquired, students could also see the final results in practice, following the analysis and interpretation of the scenarios that had been assigned.

In a learning scenario in the *Pedagogy and Didactics of Financial Accounting* course students had to be both learners and teachers, identifying and analysing the content to be taught. They also had to create interactive materials/presentations and to research which teaching methods were the most appropriate, depending on the assigned subject, in order to capture the attention of their colleagues [1].

In the *Virtual Learning Environments in Economics* course, students learned how to create interactive platforms that allowed them to conduct educational activities at a high level of performance [1]. Also, trainers from the labour market organizations taught them different teaching methods and software (Canva, Google sites), so that they would will be able to implement their lessons in the near future at high quality standards.

In the third course - *Business Communication* - students learned different theoretical notions about the communication process (techniques, channels, limits/constraints etc.) [1]. They

Co-funded by the Erasmus+ Programme of the European Union

were put in different situations: difficulties in implementing a project, identifying obstacles in a business meeting, analysing documents in order to develop the ability of critical thinking.

Considering that most of the learning scenarios in which students were involved engaged them in analysis and interpretation of information, there were significant changes in *Interpretation* skills and *Perseverance* among students as a result of the interventions of trainers from the labour market organizations. In addition, throughout the three courses students succeeded in developing certain dispositions (to a higher or lower level), such as: attentiveness, open-mindedness, intrinsic goal motivation (due to case studies that had applicability in practice). In this way, students could form a global perspective: from theoretical notions to the result in real life. Overall, and considering that the intervention of the labour market trainers had a positive impact on the development of the students' ability to think critically, it is recommended to harmonize the teaching methods/techniques used by the teachers in higher education institutions with those promoted by the labour market representatives.

Co-funded by the Erasmus+ Programme of the European Union

PART III – THINK4JOBS GUIDELINES FOR CTBACS IMPLEMENTATION

CTBACs implementation occurred in the first academic year after the Covid pandemics adjustments, in the fall and spring terms of 2021-2022. In most countries it represented the first year of face-to-face teaching since the declaration of the pandemic; the concerns about Academia's lockdown were still fresh in everyone's mind. Nonetheless, the students' engagement proved they were eager to resume classes and participate in the designed activities, even if they anticipated an increased workload compared with the one in the traditional learning process. In some context, the development of activities in groups was a positive factor when the task seemed too demanding.

Overall, the results of the piloting CTBACs indicate that regardless of the discipline in which critical thinking was addressed, the interventions were proven beneficial in terms of CT skills and dispositions transferability from curricula to apprenticeships and from the labour market to the Academia (from apprenticeships into the curricula). The close cooperation with LMO linked the labour market and the professional settings together, bringing into the Academy a new dynamic teaching approach, both parts benefiting of the use of various active methods in the courses held within the project, with a particular emphasis on discovery learning through one's own experiences [65].

Below we present a set of guidelines regarding the implementation of the CTBACs, rooted in the experience gathered during the piloting of the blended courses.

1. Explain what you are doing – Why is CT important in the labour market?

It is important to explain explicitly to the students why CT is a determinant competency in today's workforce, highly praised by the labour market in all the professions [8, 66, 67], highlighting the specificities that may result from the differences in the professional fields [68,

Co-funded by the Erasmus+ Programme of the European Union

69]. This can be achieved by bringing into the classroom the experience of the labour market, either by inviting the discussion with well positioned professionals, or allowing students to visit and assess first-hand the work of stakeholders. The use of real work situations to train CT approaches to solving problems is a crucial step during training for better professionals, with increased experience in taking informed decisions at problem solving.

Then, clarify how the instructional method designed for the CT blended curricula match the activities of professionals that daily deal with specific challenges, situations, or problems, and how those activities were planned to strengthen the students' ability to deal with them, and to solve them by themselves, therefore aiming to facilitate their introduction in the labour market. As defended by Abrami and collaborators [70] the explicit development of CT drives higher gains particularly if one is also using an immersive approach, such as it happened in the implemented CTBACs. Our results show that content-specific CT skills and dispositions were developed in students enrolled in CTBACs, which was supported by the earlier work of Abrami et al [71].

Some of the strategies used in the piloting courses related to an action-oriented approach. By focusing on the learner as a social agent, on student-centeredness, on the social or the professional context, the used interventions immersed students in cognitively challenging situations and engaged them in real-life tasks that require the activation of general competencies and CT skills. Our results also indicate that explicit instruction of CT proves to benefit students' acquisition of CT skills and dispositions. Still, in the Critical Thinking Blended Apprenticeships Curricula designed for the current interventions, explicit instruction of CT was implemented differently according to the programme/discipline, because of its specificities (in either the qualifications imposed by the course' syllabus or the operationalization of the apprenticeships). However, in most cases, the explanation of the CT conceptualization and its importance was only discussed with students at the beginning of the courses. We argue that if instructors systematically infuse it during their courses and content instruction, the students' reflection and self-regulation could be enhanced.

Data gathered in some courses seems to suggest that it is important to reinforce multiple times across a course that the CT skills and dispositions are additional (but essential) outcomes for the course. Students might have an expectation from the course that this is an ordinary course where most attention is paid to the acquisition of cognitive knowledge but not to the quality of the task performance. Students should be informed about the purposes of tasks and the learning outcomes that encompass the development of not only subject-related cognitive competencies but general soft competencies, the latter including the development of CT skills and dispositions. Students' awareness of the relevance of CT skills when learning a discipline should be raised throughout the whole course. This is in line with El Soufi and See [72] that showed that only explicit instruction of CT skills proved to have the best evidence of effectiveness.

For coherence, another issue that must be considered is the fact that if the development of CT skills has been identified as crucial for students' apprenticing, then specific CT skills and dispositions should be identified in the learning outcomes established for the course or the modules, intertwined with its content knowledge, which congruently should also be assessed during the course.

2. CT training must be a continuous and pervasive process

The students enrolled in this project acted as a selected cohort in five different disciplines [Business Informatics (Germany), Teacher Education (Greece), Veterinary Medicine (Portugal) and Business and Economics (Romania), as well as the course English as Foreign Language (Lithuania)]. They received specific training aiming at equipping them to analyse and decide on the needed action to solve a problem typical in the daily life of their professions. Although the Critical Thinking Apprenticeships Curricula were implemented in each Programme during the academic year 2021-2022 (see IO3 for more information regarding the designed courses [1]), we argue that instruction for CT should be organized as a continuous and pervasive process that should be promoted across the entire Programme. It should start at the beginning

Co-funded by the Erasmus+ Programme of the European Union

of each course in the programme plan and culminate in the final apprenticeship, which usually is placed at the end of the undergraduate studies. In that sense, students will have a constant engagement and opportunity to exercise and cultivate their CT skills and dispositions.

This would be particularly important for CT dispositions, which require internalization [55] and the acquisition of habits of mind. A curriculum explicitly orientated to CT should be adopted and applied not only in specific courses (more closely related to apprenticeships) but in every course of a discipline at the University level. In that way, students can be familiarized and constantly practice the concept. Hence, when entering apprenticeships students are expected to be aware of the CT skills and dispositions they can exploit while solving problems in real-life situations as well as maximizing the possibilities for transferring their acquired CT skills and dispositions in new contexts. At the end, time required for adaptation to labour market demands would be reduced, the stress of newly graduates reduced, while the quality of the work provided and the wellbeing of the novice professionals would greatly increase.

3. Get time to do it

The CTBACs may be viewed as a capacity-building instrument to foster CT across HE graduation programmes, designed to facilitate the integration of CT skills and dispositions into HE students and to mitigate the competencies´ gaps or mismatches reported by stakeholders and policy makers.

Nevertheless, for CTBACs to result, they request much needed time from professors and students to work on the proposed goals. The timeframe for the educational interventions must be carefully established. If more than one course in the same year of the programme is using this educational approach, then the time frame and learning outcomes should be considered together. Care must be taken to avoid building-up the workload of students (and teachers), so the engagement and gains would be expectably higher. Together with the repetition driven by the introduction of this approach across the curricula, the cumulative

Co-funded by the Erasmus+ Programme of the European Union

gains would be consistently obtained not only for CT skills, but also for CT dispositions, since attitudinal changes will require more time to be absorbed.

Careful planning is necessary to develop CTBACs, since they focus on the development of soft skills training like CT. Quality education and consistent skills improvement is affected by the time made available for learning. They require either good planning, or adequate classroom management and the ability to adjust the time on tasks to the characteristics of the group of students. Even though HE classes are usually less structured than those in other educational levels, the interventions should be carefully designed and learning activities adequately framed to maximize students' engagement and gain. Effective feedback must be prompt and clear, meaningful for the student and task at hand, and provide guidance respecting needed improvements [73]. It should also aim at correcting possible shortcomings, dispersion from the proposed goal of the instruction or falling back of schedule associated with procrastination or other self-regulatory difficulties. Therefore, feedback must be planned strategically, to manage the educator workload and a timely intervention, and to include moments for students' self-correction, concentrating teachers' feedback in key moments of the activities. It is important to plan the moments, location, and the sequences of feedback events [74].

4. Get connected to reality - Motivate students with authentic and experiential learning

The impact of authentic instruction or experiential learning alone on CT skills and dispositions acquisition was not evaluated during the current intervention. Still, the Critical Thinking Blended Apprenticeships Curricula employed instruction with case studies, real/authentic problems, and critical incidents across the different implemented courses, interweaving the theoretical and practical aspects of the topics delivered. Linking the factual and conceptual knowledge that students need to acquire with practical examples issued from daily professional life increases the level of inherent students' motivation for learning and turns them into positive experiences [64, 75] where students have the opportunity to test multiple approaches to (autonomously) try and solve problems. Therefore, the interventions should be

Co-funded by the Erasmus+ Programme of the European Union

challenging, close to the reality of the profession and allow the students some autonomy in decision-making.

CTBACs must allow students to develop CT-related behaviours they perceive as useful for any professional goals they establish for themselves. By bringing the Labour Market experiences into the training (either during apprenticeships or in the classroom) we consider CT skills and dispositions are nurtured and reinforced in students. In addition, any impact of authentic as well as experiential learning is expected to be maximized as long as students enter the apprenticeships and transfer the skills and dispositions in real-life contexts.

5. Take/accept reasonable risk

CT implies facing complex problems with uncertain solutions. Only in this kind of situation the students, the trainees and even the tutors need to face new ill-defined situations to try and solve the problem and expand their CT. If one's using memory in a certain situation, one will not use critical thinking skills to solve the problem. Therefore, it is mandatory to put students in situations where they need to ponder several different solutions for a problem and choose the one they see as the more suitable. Moreover, maybe they will be wrong. Consequently, they must start again, and again. Nevertheless, this possibility of failure is necessary to develop CT skills and dispositions. Therefore, while planning the training programs, the error must be previewed, and the risk of failure accepted.

Making errors (understood as an incorrect decision making), followed by corrective feedback, and managing failure is a strong learning experience [76, 77]. The corrective feedback must result from the analysis of the underlying reason for it to be consequential [76].

Experiencing and identifying errors foster self-regulation (namely, self-monitoring and self-assessment), the exploitation of alternative solutions and metacognition. Besides it also enhances subsequent memory retrieval, and the ability to correctly ponder subtle cues into the overall scenario, while mitigating the effects of high-and low-confidence decisions [76].

Co-funded by the Erasmus+ Programme of the European Union

Experiencing deficient decision-making in the classroom context stimulates the development of CT skills and attitudes in students and will expectably occur. Therefore, monitoring strategies and feedback cycles should be planned and explicitly incorporated within the interventional strategies to increase the gains in CTBACs courses.

6. Reflect on CT skills and dispositions changes.

As part of the feedback, or in parallel with it, reflecting on the learning process is an integral part of developing CT skills and dispositions, and is a sought attitude both in learning and in professional settings. Offering the students the opportunity to reflect on their learning process is a way for them to go beyond simply participating and be focused on the 'how' to complete their task or activity, driving them to explore why they are doing it [78]. Metacognition, a higher CT skill, involves the awareness of one's thinking or thinking about thinking. As Nappi states [79] metacognition is an essential skill that needs to be honed to recognize how one learns.

In the current study, the designed Critical Thinking Blended Apprenticeships Curricula offered different reflective opportunities. For these opportunities to be meaningful, they should be explicit. We defend that students should be offered the opportunity to reflect upon the procedure and final results of the tasks they practiced. This leads students to rethink and evaluate the strong and weak points of their performance and will foster a deeper understanding of what has been expected from them, and whether they have performed the tasks up to the required standards. This process will also engage students in the application of CT skills and will reveal that the tasks do not only require recalling information but more importantly inviting them to analyse it, apply it, and create new forms of knowledge. Teaching students how to think about their thinking or metacognition can lead students to a deeper understanding. Further, reflection as well as self-assessment can be a guided process where mentors can also be involved explicitly. At the same time, reflection among peers can be also applied through focus group discussion with a rubric of questions, where students can share

Co-funded by the Erasmus+ Programme of the European Union

their apprenticeship material and diaries as well as discuss them in a relaxed manner. This way can probably open new opportunities for CT cultivation since stressful factors such as academic performance are excluded, and therefore should be incorporated in the interventions when designing new CT-embedded curricula. This kind of group reflection can be organized as a mandatory task and blogs as well as Moodle can be utilized to this end.

Co-funded by the Erasmus+ Programme of the European Union

References

[1] Mäkiö J, Mäkiö E, Pnevmatikos D, Christodoulou P, Payan Carreira R, Georgiadou T, et al. THINK4JOBS CRITICAL THINKING CURRICULA: Critical Thinking blended apprenticeships curricula. Greece: University of Western Macedonia; 2022.

[2] Payan-Carreira R, Sacau-Fontenla A, Rebelo H, Sebastião L, Pnevmatikos D. Development and Validation of a Critical Thinking Assessment-Scale Short Form. Education Sciences2022.

[3] Quinn S, Hogan M, Dwyer C, Finn P, Fogarty E. Development and Validation of the Student-Educator Negotiated Critical Thinking Dispositions Scale (SENCTDS). Thinking Skills and Creativity. 2020;38:100710.

[4] OECD. Does Higher Education Teach Students to Think Critically?2022.

[5] Heyneman SP. International education quality. Economics of Education Review. 2004;23:441-52.

[6] Indrasiene V, Jegeleviciene V, Merfeldaite O, Penkauskiene D, Pivoriene J, Railiene A, et al. What Critical Thinking and for What? Social Welfare Interdisciplinary Approach. 2019;9:24-38.

[7] Zahner D, Van Damme D, Benjamin R, Lehrfeld J. Measuring the generic skills of higher education students and graduates: Implementation of CLA+ international. Assessing undergraduate learning in psychology: Strategies for measuring and improving student performance. Washington, DC, US: American Psychological Association; 2021. p. 219-41.

[8] Pnevmatikos D, Christodoulou P, Georgiadou T, Lithoxoidou A, Dimitriadou A, Payan Carreira R, et al. THINK4JOBS TRAINING: Critical Thinking Training Packages for Higher Education Instructors and Labour Market Tutors: University of Western Macedonia; 2021.

[9] Rebelo H, Sebastião L, Ferreira D, Payan-Carreira R. Developing Critical Thinking in Higher Education: Is There a Reason to Change? In: Reis A, Barroso J, Martins P, Jimoyiannis A, Huang RY-M, Henriques R, editors. Technology and Innovation in Learning, Teaching and Education. Cham: Springer Nature Switzerland; 2022. p. 329-41.

[10] Wilkes M, Bligh J. Evaluating educational interventions. BMJ. 1999;318:1269-72.

[11] Rear D. One size fits all? The limitations of standardised assessment in critical thinking. Assessment & Evaluation in Higher Education. 2019;44:664-75.

[12] Verburgh A, François S, Elen J, Janssen R. The Assessment of Critical Thinking Critically Assessed in Higher Education: A Validation Study of the CCTT and the HCTA. Education Research International. 2013;2013:198920.

Co-funded by the Erasmus+ Programme of the European Union

[13] Hart C, Da Costa C, D'Souza D, Kimpton A, Ljbusic J. Exploring higher education students' critical thinking skills through content analysis. Thinking Skills and Creativity. 2021;41:100877.

[14] Braun HI, Shavelson RJ, Zlatkin-Troitschanskaia O, Borowiec K. Performance Assessment of Critical Thinking: Conceptualization, Design, and Implementation. Frontiers in Education. 2020;5.

[15] Williamson DM, Xi X, Breyer FJ. A Framework for Evaluation and Use of Automated Scoring. Educational Measurement: Issues and Practice. 2012;31:2-13.

[16] Liu OL, Frankel L, Roohr KC. Assessing Critical Thinking in Higher Education: Current State and Directions for Next-Generation Assessment. ETS Research Report Series. 2014;2014:1-23.

[17] Payan-Carreira R, Cruz G, Papathanasiou IV, Fradelos E, Jiang L. The effectiveness of critical thinking instructional strategies in health professions education: a systematic review. Studies in Higher Education. 2019;44:829-43.

[18] Hyytinen H, Ursin J, Silvennoinen K, Kleemola K, Toom A. The dynamic relationship between response processes and self-regulation in critical thinking assessments. Studies in Educational Evaluation. 2021;71:101090.

[19] Kreitchmann RS, Abad FJ, Ponsoda V, Nieto MD, Morillo D. Controlling for Response Biases in Self-Report Scales: Forced-Choice vs. Psychometric Modeling of Likert Items. Frontiers in Psychology. 2019;10.

[20] Bravo MJ, Galiana L, Rodrigo MF, Navarro-Pérez JJ, Oliver A. An adaptation of the Critical Thinking Disposition Scale in Spanish youth. Thinking Skills and Creativity. 2020;38:100748.

[21] Toplak ME, West RF, Stanovich KE. Rational thinking and cognitive sophistication: Development, cognitive abilities, and thinking dispositions. Developmental Psychology. 2014;50:1037-48.

[22] Facione PA. Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction (The Delphi Report). In: Association AP, editor. Educational Resources Information Center (ERIC). Newark, DE, USA: Millbrae, CA: California Academic Press; 1990. p. 112 p.

[23] Facione PA. The Disposition Toward Critical Thinking: Its Character, Measurement, and Relationship to Critical Thinking Skill. Informal Logic. 2000;20:61 - 84.

[24] Nair G. Preliminary psychometric characteristics of the critical thinking self-assessment scale. Saskatoon: University of Saskatchewan; 2011.

Co-funded by the Erasmus+ Programme of the European Union

[25] Nair GG, Hellsten LM, Stamler LL. Accumulation of Content Validation Evidence for the Critical Thinking Self-Assessment Scale. J Nurs Meas. 2017;25:156-70.

[26] Gudmundsson E. Guidelines for translating and adapting psychological instruments. Nordic Psychology. 2009;61:29-45.

[27] Tsang S, Royse CF, Terkawi AS. Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. Saudi J Anaesth. 2017;11:S80-S9.

[28] Gerdts-Andresen T, Hansen MT, Grøndahl VA. Educational effectiveness: Validation of an instrument to measure students' critical thinking and disposition. International Journal of Instruction. 2022;25:685 - 700.

[29] Flora DB, Curran PJ. An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. Psychol Methods. 2004;9:466-91.

[30] Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal. 1999;6:1-55.

[31] Hair JF, Page M, Brunsveld N. Essentials of Business Research Methods. 4th ed. ed. New York, NY: Routledge.; 2019.

[32] Cheung GW, Rensvold RB. Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. Structural Equation Modeling: A Multidisciplinary Journal. 2002;9:233-55.

[33] Chen FF. Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. Structural Equation Modeling: A Multidisciplinary Journal. 2007;14:464-504.

[34] Muthén LK, Muthén BO. Mplus User's Guide. 7th Ed. ed. Los Angeles, CA: Muthén & Muthén; 2012.

[35] Marôco J. Análise de Equações Estruturais - Fundamentos teóricos, software & aplicações. 2nd ed. Pero Pinheiro: ReportNumber, Análise e Gestão de Informação, Ltd; 2014.

[36] Maroco J. Análise Estatística com o SPSS Statistics. 7th Ed. ed. Pero Pinheiro: ReportNumber- Análise e gestão de Informação, Ltd; 2018.

[37] Polat S. Multidimensional a nalysis of the teaching process of the critical thinking skills. Research in Social Sciences and Technology. 2020;5:134 - 57.

[38] Bensley DA, Murtagh MP. Guidelines for a Scientific Approach to Critical Thinking Assessment. Teaching of Psychology. 2011;39:5-16.

Co-funded by the Erasmus+ Programme of the European Union

[39] Lewis A, Smith D. Defining Higher Order Thinking. Theory Into Practice. 1993;32:131-7.

[40] Taber KS. The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. Research in Science Education. 2018;48:1273-96.

[41] Ku KYL. Assessing students' critical thinking performance: Urging for measurements using multi-response format. Thinking Skills and Creativity. 2009;4:70-6.

[42] Putnick DL, Bornstein MH. Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. Developmental Review. 2016;41:71-90.

[43] Moody DL, Sindre G. Evaluating the effectiveness of learning interventions: an information systems case study. European Conference on Information Systems2003.

[44] Dillman DA, Smyth JD, Christian LM. Internet, phone, mail, and mixed mode surveys: The tailored design method, 4th ed. Hoboken, NJ, US: John Wiley & Sons Inc; 2014.

[45] Nair CS, Adams P, Mertova P. Student Engagement: The Key to Improving Survey Response Rates. Quality in Higher Education. 2008;14:225-32.

[46] Friend CM, Zubek JP. The Effects Of Age On Critical Thinking Ability. Journal of Gerontology. 1958;13:407-13.

[47] Karagöl İ, Bekmezci S. Investigating Academic Achievements and Critical Thinking Dispositions of Teacher Candidates. Journal of Education and Training Studies. 2015;3:86-92.

[48] Nieto AM, Valenzuela J. A study of the internal structure of critical thinking dispositions. Inquiry: Critical Thinking Across the Disciplines. 2012;27:31-8.

[49] Mäkiö E, Mäkiö J. Teaching Critical Thinking– A Task-Based Approach: Work in Progress. In: Reis A, Barroso J, Martins P, Jimoyiannis A, Huang RY-M, Henriques R, editors. Technology and Innovation in Learning, Teaching and Education. Cham: Springer Nature Switzerland; 2022. p. 265-73.

[50] Qing Z, Ni S, Hong T. Developing critical thinking disposition by task-based learning in chemistry experiment teaching. Procedia - Social and Behavioral Sciences. 2010;2:4561-70.

[51] Halpern DF. Teaching critical thinking for transfer across domains: Disposition, skills, structure training, and metacognitive monitoring. American Psychologist. 1998;53:449-55.

[52] Burbach ME, Matkin GS, Quinn CE, Searle TP. The Impact of Preparing Agriculture Faculty to Influence Student Critical Thinking Disposition. Journal of Agricultural Education. 2012;53:1-14.

Co-funded by the Erasmus+ Programme of the European Union

[53] Giancarlo CA, Facione PA. A look across four years at the disposition toward Critical Thinking among undergraduate students. The Journal of General Education. 2001;50:29-55.

[54] Lampert N. Critical Thinking dispositions as an outcome of undergraduate education. The Journal of General Education. 2007;56:17-33.

[55] Bloch J, Spataro SE. Cultivating Critical-Thinking Dispositions Throughout the Business Curriculum. Business and Professional Communication Quarterly. 2014;77:249-65.

[56] Marin LM, Halpern DF. Pedagogy for developing critical thinking in adolescents: Explicit instruction produces greatest gains. Thinking Skills and Creativity. 2011;6:1-13.

[57] Heijltjes A, Gog TV, Paas F. Improving students' critical thinking: Empirical support for explicit instructions combined with practice. Applied Cognitive Psychology. 2014;28:518-30.

[58] Europe Co. Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume. Strasbourg: Council of Europe Publishing; 2020.

[59] Piccardo E, North B. The Action-oriented Approach: Multilingual Matters; 2019.

[60] Kriaučiūnienė R, Targamadzė V, Arcimavičienė, L. Insights into the Application of Action-oriented Approach to Language Teaching and Learning at University Level: a case of Vilnius University. International Journal of Multilingual Education. 2020;16:1-22.

[61] Colucciello ML. Relationships between critical thinking dispositions and learning styles. Journal of Professional Nursing. 1999;15:294-301.

[62] Payan-Carreira R, Silva R, Simões M, Rebelo H. Business-University Collaboration in Designing Work-Based Activities Fostering Clinical Reasoning. In: Reis A, Barroso J, Martins P, Jimoyiannis A, Huang RY-M, Henriques R, editors. Technology and Innovation in Learning, Teaching and Education. Cham: Springer Nature Switzerland; 2022. p. 342-53.

[63] Payan-Carreira R, Rebelo H, Sebastião L. Perspective Chapter: Active Learning Strategies in the Veterinary Medicine Programme under the Think4Jobs Project. In: Ortega-Sánchez D, editor. Active Learning. Rijeka: IntechOpen; 2022. p. Ch. 7.

[64] Payan-Carreira R, Sebastião L, Cristóvão AM, Rebelo H. How to Enhance Students' Self-Regulation. In: Dutton J, editor. The Psychology of Self-Regulation. Hamilton, USA: Nova Science Publishers, Inc; 2022. p. 211 -32.

[65] Dumitru D, Christodoulou P, Lithoxoidou A, Georgiadou T, Pnevmatikos D, Drămnescu AM, et al. THINK4JOBS TOOLKIT Ten work-based learning scenarios. University of Western Macedonia, Greece; 2021.

Co-funded by the Erasmus+ Programme of the European Union

[66] World Economic Forum. The future of jobs: employment, skills and workforce strategy for the Fourth Industrial Revolution. Geneva: World Economic Forum; 2016. p. vi, 157 p.

[67] OECD. Getting Skills Right: Skills for Jobs Indicators2017.

[68] Grosemans I, Coertjens L, Kyndt E. Exploring learning and fit in the transition from higher education to the labour market: A systematic review. Educational Research Review. 2017;21:67-84.

[69] Cruz G, Payan-Carreira R, Dominguez C, Silva H, Morais F. What critical thinking skills and dispositions do new graduates need for professional life? Views from Portuguese employers in different fields. Higher Education Research & Development. 2021;40:721-37.

[70] Abrami PC, Bernard RM, Borokhovski E, Wade A, Surkes MA, Tamim R, et al. Instructional interventions affecting critical thinking skills and dispositions: A stage 1 meta-analysis. Review of Educational Research. 2008;78:1102-34.

[71] Abrami PC, Bernard RM, Borokhovski E, Waddington DI, Wade CA, Persson T. Strategies for teaching students to think critically: A meta-analysis. Review of Educational Research. 2015;85:275-314.

[72] El Soufi N, See BH. Does explicit teaching of critical thinking improve critical thinking skills of English language learners in higher education? A critical review of causal evidence. Studies in Educational Evaluation. 2019;60:140-62.

[73] Shute VJ. Focus on Formative Feedback. Review of Educational Research. 2008;78:153-89.

[74] Boud D, Dawson P. What feedback literate teachers do: an empirically-derived competency framework. Assessment & Evaluation in Higher Education. 2021:1-14.

[75] Kong Y. The Role of Experiential Learning on Students' Motivation and Classroom Engagement. Frontiers in Psychology. 2021;12.

[76] Metcalfe J. Learning from Errors. Annual Review of Psychology. 2017;68:465-89.

[77] Mera Y, Rodríguez G, Marin-Garcia E. Unraveling the benefits of experiencing errors during learning: Definition, modulating factors, and explanatory theories. Psychonomic Bulletin & Review. 2022;29:753-65.

[78] McLeod SA. *Kolb - learning styles*. Simply Psychology2017. p. 8.

[79] Nappi JS. The Importance of Questioning in Developing Critical Thinking Skills. Delta Kappa Gamma Bulletin. 2017;84:34-41.

Co-funded by the Erasmus+ Programme of the European Union

Supplementary material

Supplementary table 1. The CTSAS-SF and SENCTDS questionnaires translated in local languages with hyperlinks.

CTSAS-SF & SENCTDS Scales in Project local languages

CTSAS-SF & SENCTDS Scales in Greek

CTSAS-SF & SENCTDS Scales in German

CTSAS-SF & SENCTDS Scales in Portuguese

CTSAS-SF & SENCTDS Scales in Lithuanian

CTSAS-SF & SENCTDS Scales in Romanian

Co-funded by the Erasmus+ Programme of the European Union

Supplementary table 2. CTSAS-SF items descriptive statistics.

Items	Mean	Sd.	Skew.	Kurt.	K-S test	p
1. I try to figure out the content of the problem.	5.04	.958	-.744	-.232	0.152	1.000
2. I classify data using a framework.	3.89	1.319	-.452	-.140	0.994	0.276
3. I break the complex ideas into manageable sub-ideas.	3.96	1.357	-.467	-.049	0.718	0.682
4. I observe the facial expression people use in a given situation	4.63	1.380	-1.071	.715	0.914	0.374
5. I examine the values rooted in the information presented.	4.12	1.284	-.532	-.172	0.754	0.620
6. I restate another person's statements to clarify the meaning.	3.63	1.515	-.359	-.545	0.762	0.607
7. I figure out an example, which explains the concept /opinion.	4.53	1.097	-.785	.550	0.601	0.863
8. I clarify my thoughts by explaining to someone else.	4.29	1.348	-.803	.203	0.864	0.445
9. I seek clarification of the meanings of another's opinion or points of view.	4.23	1.185	-.483	-.196	0.718	0.682
10. I examine the similarities and differences among the opinions posed for a given problem.	4.23	1.166	-.742	.765	0.518	0.951
11. I examine the interrelationships among concepts or opinions posed.	3.84	1.222	-.364	.101	0.629	0.823
12. I look for supporting reasons when examining opinions.	4.44	1.174	-.692	.436	0.640	0.808
13. I look for relevant information to answer the question at issue.	4.62	1.147	-.855	.657	0.651	0.790
14. I examine the proposals for solving a given problem.	4.65	1.089	-.626	-.100	0.260	1.000
15. I ask questions in order to seek evidence to support or refute the author's claim.	4.09	1.341	-.566	-.084	1.041	0.229
16. I figure out if author's arguments include both for and against the claim.	3.97	1.316	-.433	-.229	1.044	0.226
17. I figure out unstated assumptions in one's reasoning for a claim.	3.63	1.289	-.287	-.190	0.723	0.673
18. I look for the overall structure of the argument.	3.99	1.332	-.580	.136	0.864	0.444
19. I figure out the process of reasoning for an argument.	4.02	1.306	-.578	.253	0.381	0.999

Co-funded by the Erasmus+ Programme of the European Union

20. I figure out the assumptions implicit in the author's reasoning.	3.73	1.275	-.436	-.032	0.828	0.500
21. I assess the contextual relevance of an opinion or claim posed.	4.00	1.192	-.493	.387	0.810	0.528
22. I seek the accuracy of the evidence supporting a given judgment.	4.18	1.283	-.693	.306	0.858	0.453
23. I assess the chances of success or failure in using a premise to conclude an argument.	4.08	1.344	-.599	-.007	1.120	0.163
24. I examine the logical strength of the underlying reason in an argument.	4.06	1.295	-.464	-.030	0.919	0.367
25. I search for new data to confirm or refute a given claim	4.15	1.288	-.644	.142	0.708	0.698
26. I search for additional information that might support or weaken an argument.	4.34	1.195	-.520	-.206	0.435	0.992
27. I examine the logical reasoning of an objection to a claim.	4.17	1.310	-.552	.025	0.883	0.417
28. I seek useful information to refute an argument when supported by unsure reasons.	4.37	1.186	-.655	.478	0.314	1.000
29. I collect evidence supporting the availability of information to back up opinions.	4.21	1.317	-.771	.585	0.794	0.554
30. I seek for evidence / information before accepting a solution.	4.49	1.241	-.729	.176	0.355	1.000
31. I figure out alternate hypotheses / questions, when I need to solve a problem.	4.21	1.311	-.645	.166	1.042	0.228
32. Given a problem to solve, I develop a set of options for solving the problem.	4.33	1.255	-.685	.234	0.683	0.739
33. I systematically analyse the problem using multiple sources of information to draw inferences.	4.11	1.381	-.596	-.103	0.325	1.000
34. I figure out the merits and demerits of a solution while prioritizing from alternatives for making decisions.	4.01	1.320	-.455	-.130	0.812	0.525
35. I identify the consequences of various options to solving a problem.	4.36	1.208	-.558	-.009	0.625	0.830
36. I arrive at conclusions that are supported with strong evidence.	4.30	1.164	-.328	-.484	0.490	0.970

Co-funded by the Erasmus+ Programme of the European Union

37. I use both deductive and inductive reasoning to interpret information.	4.00	1.330	-.419	-.259	0.766	0.600
38. I analyse my thinking before jumping to conclusions.	4.39	1.335	-.710	.065	0.437	0.991
39. I confidently reject an alternative solution when it lacks evidence.	3.89	1.417	-.312	-.587	0.541	0.932
40. I figure out the pros and cons of a solution before accepting it.	4.64	1.175	-.721	.216	0.710	0.695
41. I can describe the results of a problem using inferential evidence.	3.78	1.206	-.269	.068	0.701	0.709
42. I can logically present results to address a given problem.	4.18	1.138	-.425	.111	1.533	0.018
43. I state my choice of using a particular method to solve the problem.	4.03	1.277	-.530	.164	0.305	1.000
44. I can explain a key concept to clarify my thinking.	4.10	1.246	-.408	-.141	0.585	0.883
45. I write essays with adequate arguments supported with reasons for a given policy or situation.	3.13	1.734	-.208	-.966	0.833	0.492
46. I anticipate reasonable criticisms one might raise against one's view points.	3.92	1.319	-.438	-.340	0.730	0.661
47. I respond to reasonable criticisms one might raise against one's view points.	3.82	1.292	-.456	-.055	1.772	0.004
48. I clearly articulate evidence for my own view points.	4.22	1.159	-.353	-.283	0.195	1.000
49. I present more evidence or counter evidence for another's points of view.	3.61	1.338	-.258	-.540	0.664	0.770
50. I provide reasons for rejecting another's claim.	4.04	1.400	-.535	-.309	1.255	0.086
51. I reflect on my opinions and reasons to ensure my premises are correct.	4.43	1.136	-.442	-.421	0.540	0.932
52. I review sources of information to ensure important information is not overlooked.	4.26	1.317	-.628	-.074	1.009	0.260
53. I examine and consider ideas and viewpoints even when others do not agree.	4.20	1.156	-.380	-.235	0.174	1.000

Co-funded by the Erasmus+ Programme of the European Union

54. I examine my values, thoughts / beliefs based on reasons and evidence.	4.41	1.159	-.455	-.151	0.143	1.000
55. I continuously assess my targets and work towards achieving them.	4.46	1.182	-.472	-.367	0.354	1.000
56. I review my reasons and reasoning process in coming to a given conclusion.	4.18	1.187	-.349	-.236	0.415	0.995
57. I analyse areas of consistencies and inconsistencies in my thinking.	4.01	1.294	-.448	-.192	0.926	0.358
58. I willingly revise my work to correct my opinions and beliefs.	4.27	1.263	-.457	-.172	0.663	0.772
59. I continually revise and rethink strategies to improve my thinking.	4.34	1.280	-.601	-.073	0.683	0.739
60. I reflect on my thinking to improve the quality of my judgment.	4.53	1.187	-.805	.752	0.235	1.000

Co-funded by the Erasmus+ Programme of the European Union

Supplementary table 3. SENCTDS items descriptive statistics.

Items	Mean	Sd.	Skew.	Kurt.	K-S test	p
1. When a theory, interpretation, or conclusion is presented to me, I try to decide if there is good supporting evidence.	5.62	1.070	-.874	1.125	.613	.847
2. When faced with a decision, I seek as much information as possible.	5.85	1.130	-1.021	.692	.934	.347
3. I try to gather as much information about a topic before I draw a conclusion about it.	5.82	1.133	-.931	.581	.562	.911
4. I find that I'm easily distracted when thinking about a task.	3.83	1.724	.049	-1.042	.900	.393
5. I find it hard to concentrate when thinking about problems.	3.90	1.827	.022	-1.133	1.179	.124
6. I often miss out on important information because I'm thinking of other things.	3.91	1.780	-.070	-1.057	1.370	**.047**
7. I often daydream when learning a new topic.	3.94	1.771	-.016	-.994	.462	.983
8. Thinking is not about 'being flexible', it's about 'being right'.	5.02	1.802	-.628	-.644	.293	1.000
9. Being open-minded about different worldviews is less important than people think.	5.52	1.702	-1.087	.134	.787	.566
10. When attempting to solve complex problems, it's better to give up fast, if you cannot reach a solution so as to not waste time.	5.46	1.684	-1.053	.106	.778	.580
11. I know what I think and believe so it's not important to dwell on it any further.	4.92	1.640	-.562	-.625	.671	.759
12. I like to make lists of things I need to do and thoughts I may have.	5.06	1.683	-.759	-.214	1.902	**.001**
13. I take notes so I can organize my thoughts.	5.19	1.653	-.900	.046	1.891	**.002**
14. I make simple charts, diagrams or tables to help me organize large amounts of information.	4.58	1.795	-.479	-.775	1.598	**.012**
15. I persevere with a task even when it is very difficult	5.40	1.256	-.561	-.208	0.339	1.000
16. Frustration does not stop me from finishing what needs to be done.	5.08	1.592	-.605	-.511	0.569	.903

Co-funded by the Erasmus+ Programme of the European Union

17. I find it desirable to keep going even if it is sometimes hard.	5.71	1.276	-1.110	1.157	0.653	.787
18. I enjoy information that challenges me to think.	5.50	1.247	-.724	.110	0.984	.287
19. I look forward to learning challenging things.	5.45	1.346	-.793	.293	1.436	**.032**
20. Completing difficult tasks is fun for me.	4.87	1.571	-.485	-.453	1.290	.072
21. Even if material is difficult to comprehend, I enjoy dealing with information that arouses my curiosity.	5.47	1.303	-.770	.290	0.703	.707

Co-funded by the Erasmus+ Programme of the European Union

Supplementary table 4: Items' loadings in CTBACS_SF

Item	*Interpretation*	*Analysis*	*Evaluation*	*Inference*	*Explanation*	*Self-Regulation*
1. I try to figure out the content of the problem.	*0.662*					
2. I classify data using a framework.	*0.661*					
3. I break the complex ideas into manageable sub-ideas.	*0.633*					
4. I observe the facial expression people use in a given situation	*0.386*					
5. I examine the values rooted in the information presented.	*0.654*					
6. I restate another person's statements to clarify the meaning.	*0.499*					
7. I figure out an example which explains the concept /opinion.	*0.594*					
8. I clarify my thoughts by explaining to someone else.	*0.422*					
9. I seek clarification of the meanings of another's opinion or points of view.	*0.536*					
10. I examine the similarities and differences among the opinions posed for a given problem.		*0.614*				
11. I examine the interrelationships among concepts or opinions posed.		*0.734*				
12. I look for supporting reasons when examining opinions.		*0.671*				
13. I look for relevant information to answer the question at issue.		*0.650*				
14. I examine the proposals for solving a given problem.		*0.701*				
15. I ask questions in order to seek evidence to support or refute the author's claim.		*0.666*				
16. I figure out if author's arguments include both for and against the claim.		*0.670*				

Co-funded by the Erasmus+ Programme of the European Union

17. I figure out unstated assumptions in one's reasoning for a claim.	*0.619*		
18. I look for the overall structure of the argument.	*0.707*		
19. I figure out the process of reasoning for an argument.	*0.772*		
20. I figure out the assumptions implicit in the author's reasoning.	*0.745*		
21. I assess the contextual relevance of an opinion or claim posed.		*0.723*	
22. I seek the accuracy of the evidence supporting a given judgment.		*0.735*	
23. I assess the chances of success or failure in using a premise to conclude an argument.		*0.702*	
24. I examine the logical strength of the underlying reason in an argument.		*0.725*	
25. I search for new data to confirm or refute a given claim		*0.674*	
26. I search for additional information that might support or weaken an argument.		*0.732*	
27. I examine the logical reasoning of an objection to a claim.		*0.761*	
28. I seek useful information to refute an argument when supported by unsure reasons.			*0.717*
29. I collect evidence supporting the availability of information to back up opinions.			*0.740*
30. I seek for evidence / information before accepting a solution.			*0.691*
31. I figure out alternate hypotheses / questions, when I need to solve a problem.			*0.734*
32. Given a problem to solve, I develop a set of options for solving the problem.			*0.710*

Co-funded by the Erasmus+ Programme of the European Union

33. I systematically analyse the problem using multiple sources of information to draw inferences.	*0.738*	
34. I figure out the merits and demerits of a solution while prioritizing from alternatives for making decisions.	*0.742*	
35. I identify the consequences of various options to solving a problem.	*0.704*	
36. I arrive at conclusions that are supported with strong evidence.	*0.756*	
37. I use both deductive and inductive reasoning to interpret information.	*0.696*	
38. I analyse my thinking before jumping to conclusions.	*0.636*	
39. I confidently reject an alternative solution when it lacks evidence.	*0.470*	
40. I figure out the pros and cons of a solution before accepting it.	*0.656*	
41. I can describe the results of a problem using inferential evidence.		*0.745*
42. I can logically present results to address a given problem.		*0.749*
43. I state my choice of using a particular method to solve the problem.		*0.672*
44. I can explain a key concept to clarify my thinking.		*0.740*
45. I write essays with adequate arguments supported with reasons for a given policy or situation.		*0.511*
46. I anticipate reasonable criticisms one might raise against one's view points.		*0.606*
47. I respond to reasonable criticisms one might raise against one's view points.		*0.650*
48. I clearly articulate evidence for my own view points.		*0.720*

Co-funded by the Erasmus+ Programme of the European Union

49. I present more evidence or counter evidence for another's points of view.	*0.573*	
50. I provide reasons for rejecting another's claim.	*0.536*	
51. I reflect on my opinions and reasons to ensure my premises are correct.		*0.719*
52. I review sources of information to ensure important information is not overlooked.		*0.785*
53. I examine and consider ideas and viewpoints even when others do not agree.		*0.705*
54. I examine my values, thoughts / beliefs based on reasons and evidence.		*0.756*
55. I continuously assess my targets and work towards achieving them.		*0.673*
56. I review my reasons and reasoning process in coming to a given conclusion.		*0.728*
57. I analyse areas of consistencies and inconsistencies in my thinking.		*0.737*
58. I willingly revise my work to correct my opinions and beliefs.		*0.750*
59. I continually revise and rethink strategies to improve my thinking.		*0.786*
60. I reflect on my thinking to improve the quality of my judgment.		*0.763*

Co-funded by the Erasmus+ Programme of the European Union

Supplementary table 5: Items' loadings in SENCTDS

Item	*Reflection*	*Attentiveness*	*Open-mindedness*	*Organization*	*Perseverance*	*Intrinsic Goal Motivation*
1. When a theory, interpretation, or conclusion is presented to me, I try to decide if there is good supporting evidence.	.755					
2. When faced with a decision, I seek as much information as possible.	.809					
3. I try to gather as much information about a topic before I draw a conclusion about it.	.834					
4. I find that I'm easily distracted when thinking about a task.		.761				
5. I find it hard to concentrate when thinking about problems.		.831				
6. I often miss out on important information because I'm thinking of other things.		.863				
7. I often daydream when learning a new topic.		.744				
8. Thinking is not about 'being flexible', it's about 'being right'.			.659			
9. Being open-minded about different worldviews is less important than people think.			.710			
10. When attempting to solve complex problems, it's better to give up fast, if you cannot reach a solution so as to not waste time.			.797			
11. I know what I think and believe so it's not important to dwell on it any further.			.694			

12. I like to make lists of things I need to do and thoughts I may have.	.720			
13. I take notes so I can organize my thoughts.	.908			
14. I make simple charts, diagrams or tables to help me organize large amounts of information.	.723			
15. I persevere with a task even when it is very difficult		.845		
16. Frustration does not stop me from finishing what needs to be done.		.735		
17. I find it desirable to keep going even if it is sometimes hard.		.819		
18. I enjoy information that challenges me to think.				.816
19. I look forward to learning challenging things.				.869
20. Completing difficult tasks is fun for me.				.698
21. Even if material is difficult to comprehend, I enjoy dealing with information that arouses my curiosity.				.796

Funding & Acknowledgements

This work has been supported by the “Critical Thinking for Successful Jobs - Think4Jobs” Project, with the reference number 2020-1-EL01-KA203-078797, funded by the European Commission/EACEA, through the ERASMUS Programme. We want to thank the different Higher Education Instructors, Higher Education Students, Labour market Tutors and Employees across the five European countries involved in the Project and in the production of IO4. In addition, we acknowledge the effort of the Think4Jobs Quality Assurance and Evaluation Committee for the improvement of the current document. Finally, we would like to thank the External Evaluation and Quality Committee/Steering Board and particularly Caroline Dominguez (University of Trásos Montes and Alto Douro) for her review of the document.

Co-funded by the Erasmus+ Programme of the European Union