



**Think4Jobs**  
Critical Thinking for Successful Jobs

# ***THINK4JOBS***

## ***Leitfaden:***

***Ein Protokoll für den  
Transfer von kritischem  
Denken aus den Lehrplänen  
in den Arbeitsmarkt***

# ***THINK4JOBS***

## ***Leitfaden:***

***Ein Protokoll für den Transfer von  
kritischem Denken aus den  
Lehrplänen in den Arbeitsmarkt***

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



## Technische and Katalogisierungsdaten

Umschlag-Layout: Entworfen in Canva

Datum der Veröffentlichung: 2023

Recommended Citation: Payan Carreira, R., Rebelo, H., Sebastião, L., Sacau, A., Ferreira, D., Simões, M., Pnevmatikos, D., Christodoulou, P., Lithoxidou, 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

Finanzierung: Diese Arbeit wurde durch das Projekt "Critical Thinking for Successful Jobs - Think4Jobs" mit der Referenznummer 2020-1-EL01-KA203-078797 unterstützt, das von der Europäischen Kommission/EACEA im Rahmen des ERASMUS+ Programms finanziert wird.

Haftungsausschluss: "Die Unterstützung der Europäischen Kommission für die Erstellung dieser Veröffentlichung stellt keine Billigung des Inhalts dar, der ausschließlich die Ansichten der Autoren wiedergibt, und die Kommission kann nicht für die Verwendung der darin enthaltenen Informationen verantwortlich gemacht werden."

## Autoren

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. Lithoxidou 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. Kriauciūnienė 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

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)

## Inhaltsverzeichnis

Autoren	3
Zusammenfassung und wichtigste Ergebnisse	7
Einführung	12
Part I –Validierung der Instrumente, die zur Bewertung der CT-Fähigkeiten und CT-Dispositionen der Studierenden eingesetzt werden	15
1. Warum und wie sollte man die CT-Veränderungen bei Studierenden nach der Einführung von CTBAC bewerten?	15
2. Methodologie	17
CTSAS-SF, das Instrument zur Bewertung der CT-Fähigkeiten im Think4Jobs-Projekt	18
SENCTDS, das Instrument zur Bewertung der CT-Dispositionen im Think4Jobs-Projekt	20
Übersetzung der Instrumente zur Bewertung der CT-Fähigkeiten und CT-Dispositionen	20
Datenerhebung und Analyse	21
3. Ergebnisse - Validierung der Skalen für CT-Fähigkeiten und CT-Dispositionen	24
Deskriptive Analyse der Items in den Skalen CTSAS-SF und SENCTDS	24
Konfirmatorische Faktorenanalyse (CFA) und Zuverlässigkeit	26
Mehrgruppeninvarianz für Geschlecht	31
Multigruppeninvarianz für Länder	34
4. Diskussion über den Validierungsprozess des Instruments	40
CTSAS Validierung	41
SENCTDS Validierung	43
Teil II – Disziplinübergreifende Analyse der CTBACs Implementation	44
1. Methodologie	46
Teilnehmer	47
Charakterisierung der Experimentalgruppe	48

Charakterisierung der Kontrollgruppe	50
Datenanalyse	50
Statistische Analyse	52
2. Ergebnisse	53
Prüfung auf Unterschiede bei den Ausgangswerten in der Versuchsgruppe	53
Beziehung zwischen Alter und CT-Fähigkeiten und -Dispositionen	54
Einfacher Gesamtvergleich vor und nach dem Test für die Experimentalgruppe	55
Fächerübergreifende Vergleiche für die Experimentalgruppe	56
Veränderungen bei den CT-Fähigkeiten und CT-Dispositionen pro Disziplin und Land	59
Deutschland	59
Griechenland	61
Litauen	66
Portugal	68
Rumänien	73
3. Diskussion	76
CTBAC-assoziierte Verbesserungen der CT-Fähigkeiten und CT-Dispositionen	80
Deutschland	81
Griechenland	83
Litauen	88
Portugal	92
Rumänien	95
Teil III – THINK4JOBS-Leitlinien für die Umsetzung von CTBACs	97
1. Erklären Sie, was Sie tun - Warum ist CT auf dem Arbeitsmarkt wichtig?	98
2. CT muss kontinuierlich und durchgängig unterrichtet werden	100
3. Zeit für die Entwicklung von CT einplanen	101

4. Get connected to reality - Motivate students with authentic and experiential learning	102
5. Angemessenes Risiko eingehen/akzeptieren	103
6. Reflektieren Sie die Veränderungen der CT-Fähigkeiten und CT-Dispositionen	104
Literatur	106
Ergänzendes Material	113
Funding & Acknowledgements	126

## Zusammenfassung und wichtigste Ergebnisse

Der intellektuelle Output 4 (IO4) berichtet über die Ergebnisse der Implementierung der Critical Thinking (CT) Blended Apprenticeships Curricula (CTBAC), die im dritten intellektuellen Output IO3 [1] beschrieben wurden, und erörtert die festgestellten Zuwächse bei den Fähigkeiten und Dispositionen zum kritischen Denken der Studierenden, die an den Pilotaktivitäten teilnahmen. Eine inter- und intradisziplinäre Analyse, die sich aus dem Vergleich zwischen den Ergebnissen vor und am Ende der CTBAC-Pilotkurse ergibt, unterstützt die von dem Projekt vorgeschlagenen Empfehlungen für die Umsetzung der Critical Thinking Blended Apprenticeships Curricula, die in Teil III dieses Berichts zusammengefasst sind: "THINK4JOBS-Leitlinien für den CT-Transfer von Lehrplänen in die Berufsausbildung".

Die Universität von Évora (UÉvora), Portugal, war der Partner, der die Durchführung von IO4 leitete. Die Ziele von IO4 wurden wie folgt definiert:

1. Bewertung der Veränderungen in den Fähigkeiten und Dispositionen zum kritischen Denken (CT) im Zusammenhang mit der Umsetzung von 12 CT Lehrplänen, wie sie von der Universität-Wirtschaftspartnerschaft für die Fächer Wirtschaftsinformatik, Lehrerbildung, Veterinärmedizin und Wirtschaftswissenschaften und den Kurs Englisch als Fremdsprache entwickelt wurden;
2. Nutzung dieser Daten für eine fächerübergreifende Analyse;

3. Präsentation der "THINK4JOBS-Leitlinien für den CT-Transfer von Lehrplänen in die Berufsausbildung".

Die Umsetzung der CTBACs stand am Anfang dieser Ziele. Die Umsetzung dieser Curricula wurde durch die enge Zusammenarbeit zwischen Hochschullehrern und Tutoren der Arbeitsmarktorganisation bei der Erstellung der Szenarien erleichtert. Die Moodle-Plattform wurde als Lernschnittstelle für die CTBACs verwendet. Die Umsetzung der neuen Lehrpläne erfolgte im Herbst- und Frühjahrssemester des Studienjahres 2021/2022. Obwohl die CTBACs im akademischen Jahr 2022/23 wiederholt werden, wird dieses Jahr in der hier vorgestellten Analyse nicht berücksichtigt.

Insgesamt nahmen 609 Studierende an den Pilotaktivitäten teil, eine größere Zahl als ursprünglich im Projektantrag vorgesehen (150 Studierende) (Tabelle 1). Dennoch beantworteten nicht alle teilnehmenden Studierenden die Fragebögen. Die Befragten, die die Fragebögen vor und nach dem letzten Test ausfüllten, stellten 54 % der Studierenden dar, die an den Aktivitäten teilnahmen. Es war schwierig, die ursprünglich vorgesehene Zahl der Kontrollgruppe (Studierende, die nicht an den Pilotkursen teilnahmen) zu erreichen, da die meisten Kurse nicht in zwei verschiedenen Semestern angeboten wurden und die Studierenden nicht akzeptierten, von Aktivitäten ausgeschlossen zu werden, von denen sie glaubten, dass sie einen Vorteil für ihren Erfolg bringen könnten. Nur dem griechischen und dem portugiesischen Partner gelang es, eine Kontrollgruppe zu organisieren, auch wenn die portugiesische Gruppe nicht sehr groß war.

Tabelle 1 – Pilot-CT Kurse, die im akademischen Jahr 2021/2022 eingeführt wurden.

Land	Programm/Disziplin	Module	Anzahl der in CTBACs eingeschriebenen Studierenden	Anzahl der Studierenden in Kontrollgruppen
Deutschland	Wirtschaftsinformatik	Design Patterns	14	--
		Innovation Management	10	--
		Economic Aspects of Industrial Digitalization	10	--
		Wissenschaftliches Seminar	10	--
Griechenland	Pädagogik	Didaktik der Biologie	83	--
		Didaktik der Naturwissenschaft	61	84
		Didaktik der Umweltkunde	12	--

Litauen	Internationale Beziehungen und Politikwissenschaft	Englisch als Fremdsprache	61	
Portugal	Veterinärmedizin	Bildgebung	78	--
		Deontologie	56	--
		Gynäkologie, Andrologie und Geburtshilfe	71	--
		Curriculare Praktika	--	12
Rumänien	Betriebs- und Volkswirtschaft	Wirtschaftskommunikation	69	--
		Pädagogik und Didaktik der	48	--
		Finanzbuchhaltung		
		Virtuelle Lernumgebungen in der Wirtschaftswissenschaft	26	--
<b>Total</b>			<b>609</b>	<b>96</b>

Der Zuwachs an Fähigkeiten und Dispositionen zum kritischen Denken der Studierenden nach der Implementierung der CTBACs wurde mit einem neuen Instrument bewertet, das zwei Fragebögen zusammenfasst (eine Kurzform des CTSAS-Fragebogens von Nair, der speziell für diesen Zweck entwickelt wurde [2], und das SENCTDS-Instrument [3]), das den Studierenden vor, während und nach den CT Interventionen vorgelegt wurde. Die Version in der Originalsprache des Instruments (Englisch) wurde ins Deutsche, Griechische, Rumänische und Portugiesische übersetzt, um von den Partnern nach eigenem Ermessen verwendet zu werden.

Die vorläufige unabhängige Validierung der beiden Fragebögen zeigte, dass sie ein starkes Instrument mit guten Anpassungsindizes und einer hohen internen Konsistenz darstellen. Die Invarianzanalyse bestätigte, dass sowohl der Teil des Instruments, der die CT Fähigkeiten bewertet, als auch die CT Dispositionen zwischen den Ländern stabil blieben, was die Qualität des Instruments und der verwendeten übersetzten Versionen (Deutsch, Griechisch, Rumänisch und Portugiesisch vs. Englisch) unterstützt. Diese Versionen stellen zusätzliche Mehrwertproduktionen aus dem Projekt dar.

Die fächerübergreifende Analyse belegt die mit der Umsetzung der gemischten Curricula verbundenen Gewinne. Es wurden einige grundlegende Unterschiede in Bezug auf Länder, Alter und Geschlecht in der Gesamtpopulation für bestimmte CT Fähigkeiten und Dispositionen festgestellt, die möglicherweise den unterschiedlichen Hintergrund der Studierenden der verschiedenen am Projekt beteiligten Disziplinen/Studiengänge widerspiegeln [Deutschland (Wirtschaftsinformatik), Griechenland (Lehrerausbildung),

Litauen (Studiengang Englisch als Fremdsprache), Rumänien (Wirtschaftswissenschaften) und Portugal (Veterinärmedizin)].

Die CTBAC-bedingten Zuwächse waren bei den Fähigkeiten deutlicher als bei den Dispositionen (ca. zehn Punkte gegenüber einem Punkt bei den jeweiligen integrierten Ergebnissen), was möglicherweise auf die Tatsache zurückzuführen ist, dass es schwieriger ist, Einstellungen (Dispositionen) bei kurzzeitigen Interventionen zu ändern. Auch wenn Kontrollgruppen nicht in allen Ländern möglich waren, es hat sich gezeigt, dass Studierende, die an CTBACs teilgenommen haben, höhere Zuwächse bei den integrierten CT-Fähigkeiten und Dispositionen erzielten als die Kontrollgruppe, insbesondere bei den Fähigkeiten *Evaluation*, *Inferenz* und *Erklärung* sowie bei der Disposition *Aufmerksamkeit*.

Ausgehend von der Interpretation der Ergebnisse, die bei der Umsetzung der Lehrpläne gesammelt wurden, wurden die Leitlinien für die Umsetzung der CTBACs in folgenden Schritten erstellt:

1. Erläutern Sie, was Sie tun - Erklären Sie den Studierenden, warum kritisches Denken in der heutigen Arbeitswelt eine entscheidende Kompetenz ist und wie es in einem bestimmten Beruf verstanden und geschätzt wird. Beziehen Sie die Entwicklung von kritischem Denken ausdrücklich in die Ergebnisse Ihres Kurses ein.
2. Die Vermittlung des kritischen Denkens muss ein kontinuierlicher und allumfassender Prozess sein. - Um erfolgreich zu sein, sind Anstrengungen in allen Fachbereichen erforderlich, um die CT-Fähigkeiten und CT-Dispositionen der Studierenden zu fördern; die Fähigkeiten müssen trainiert und die Dispositionen verinnerlicht werden, so dass Zeit und gemeinsame Anstrengungen auf der Ebene der Fachbereiche erforderlich sind, um größere und beständige Fortschritte oder positive Veränderungen zu erzielen.
3. Nehmen Sie sich Zeit - wie bereits erwähnt, ist auch auf Kursebene Zeit erforderlich, um an den gesetzten Zielen zu arbeiten. Sowohl die Studierende als auch die Lehrende müssen die Aktivitäten so planen, dass sie Zeit haben, sich vorzubereiten, zu entwickeln und Feedback zu den Leistungen der Studierenden zu geben oder zu erhalten, damit die Intervention zu einem sinnvollen Lernen führt.

4. Realitätsbezug - Durch die Darstellung von Fällen, die aus dem Berufsalltag stammen, steigt die Motivation der Studierenden und sie nehmen die Lernerfahrungen positiver wahr.
5. Akzeptieren Sie ein angemessenes Risiko - die Entwicklung des kritischen Denkens kann davon profitieren, den Studierenden komplexe Probleme mit unsicheren Lösungen zu präsentieren, bei denen die Studierende stolpern dürfen, da ein Fehlschlag (eine falsche Entscheidung) in einem sicheren Umfeld, in dem die Studierende über die Prämissen nachdenken können, die zu dem Fehler geführt haben, und das korrigierende Feedback von Dozenten ein starkes Lernmittel sind.
6. Reflektieren Sie die Entwicklung von CT-Fähigkeiten und CT-Dispositionen - bieten Sie den Studierenden die Möglichkeit, über die Veränderungen in ihrer Denkweise oder die von ihnen entwickelten Dispositionen nachzudenken, um die Wirkung der Lerninterventionen zu verstärken und ein reflektiertes Denken über die eigenen Erfahrungen besser zu fördern.

## Einführung

Laut OCDE klafft eine zunehmende Lücke zwischen den von den Hochschulen vermittelten Qualifikationen und den von den Arbeitsmarktorganisationen (LMO) geforderten allgemeinen Fähigkeiten des 21. Jahrhunderts, insbesondere in Bezug auf Lese- und Schreibfähigkeiten und Fähigkeiten zum kritischen Denken, die Problemlösung, analytisches Denken und Kommunikationskompetenz umfassen [4]. Nach Ansicht der Arbeitgeber fördert CT die Bedingungen für eine Person, sich kontinuierlich zu verbessern, um sich schnell an den organisatorischen Wandel anzupassen [5]. Gleichzeitig sind die Arbeitgeber der Ansicht, dass kritisches Denken die Mitarbeiter in die Lage versetzt, sich schneller an die Herausforderungen eines sich verändernden Umfelds anzupassen und die besten Lösungen für das Unternehmen, die Kunden und sich selbst zu finden [6]. Wenn man die Rolle der Hochschulen in der formalen Berufsausbildung und die Menge der Absolventen, die jedes Jahr auf den Arbeitsmarkt kommen, bedenkt, kann diese Aussage beunruhigend sein, da man verstehen kann, dass ein Hochschulabschluss heutzutage den Absolventen nicht unbedingt höhere Kompetenzen für einen bestimmten Beruf garantiert. Zum Teil kann dieses Problem aus einem Kontext resultieren, in dem sich die Nachfrage nach Qualifikationen schnell verändert, um dem konstant hohen Tempo zu entsprechen, das durch das schnelle Wachstum des technologischen und wissenschaftlichen Wissens angetrieben wird.

This awareness drove the higher education reform, which shifted the learning approach from a lecture format to a student-centred one, requesting the active participation of students in the learning process, requiring the students to transpose the acquired knowledge into the solution of new or different problematic situations [7], to support decision-making and search new forms to solve a situation or address a problem.

Diese Asymmetrien in den Kompetenzen der frischgebackenen Hochschulabsolventen standen im Mittelpunkt des Think4Jobs-Projekts, das dieses Problem mit Hilfe eines kollaborativen Ansatzes angegangen ist, bei dem Hochschulen und LMO bei der Entwicklung von Critical Thinking Blended Apprenticeships Curricula (CTBAC) zusammenarbeiten, die auf die Entwicklung von Fähigkeiten und Dispositionen im Bereich des kritischen Denkens (CT)

abzielen und gemischte Interventionen in einigen Kursen eines bestimmten Fachbereichs umsetzen [8, 9].

Die CTBAC wurden mit Hilfe der E-Learning-Plattform Moodle umgesetzt. Diese Plattform ermöglichte es den Lehrkräften und den Beteiligten, die Interventionen in jedem Kurs separat einzusetzen und die Teilnahme der in den Kursen eingeschriebenen Lernenden zu unterstützen. Darüber hinaus dient die Plattform als Speicher für das in den Kursen bereitgestellte Material und zeichnet die Ergebnisse (Noten) der Studierenden in den Lernaktivitäten auf, die am Ende zu den Endnoten beitragen können. Die Umsetzung der Lernaktivitäten erfolgte nach dem in IO3 [1] vorgeschlagenen Design. Eine wichtige Frage, die im Rahmen des Projekts untersucht werden soll, ist, ob das neue Unterrichtsdesign am Ende der Pilotkurse zur Verbesserung der CT-Fähigkeiten der Studierenden beigetragen hat, wie es im Kursdesign vorgesehen war, und um den relativen Fortschritt oder den kurzfristigen "Lerngewinn" in Bezug auf die CT-Fähigkeiten und CT-Dispositionen zu ermitteln. Um die notwendigen Informationen zu sammeln, wurde ein Vortest/Nachtest-Ansatz gewählt, wobei als Mittel zur Datenerhebung ein Instrument verwendet wurde, das zwei Fragebögen zusammenfasste, von denen einer die CT-Fähigkeiten und der andere die CT-Dispositionen betraf.

Dieser intellektuelle Output - IO4 – hat folgende Ziele gesetzt:

1. Identifizierung von Veränderungen in den CT-Fähigkeiten und CT-Dispositionen bei Studierenden, die an den CT-Lehrplänen teilnehmen, basierend auf dem Vergleich der in jedem Kurs/jeder Disziplin erzielten Ergebnisse bei einem Vor- und Nachtest;
2. Bewertung der Ergebnisse der Implementierung der neuen Lerninterventionen und deren Diskussion, um den Bedarf an möglichen Änderungen der Lernszenarien zu identifizieren;
3. Vorstellung der "THINK4JOBS-Leitlinien für den Transfer von CT-Lehrplänen in die Berufsausbildung".

Die Präsentation der Informationen in IO4 ist in drei Hauptabschnitte unterteilt, beginnend mit der Auswahl und Validierung der Fragebögen, die für die Bewertung der CT-Fähigkeiten

und CT-Dispositionen verwendet wurden, gefolgt von der Präsentation der inter- und intradisziplinären Datenanalyse und der Interpretation der Ergebnisse, die die Empfehlungen unterstützen, die als THINK4JOBS-Leitlinien für die Umsetzung der CTBACs vorgestellt wurden.

## TEIL I –Validierung der Instrumente, die zur Bewertung der CT-Fähigkeiten und CT-Dispositionen der Studierenden eingesetzt werden

### 1. Warum und wie sollte man die CT-Veränderungen bei Studierenden nach der Einführung von CTBAC bewerten?

Die Evaluierung von Bildungsmaßnahmen ist von entscheidender Bedeutung, um den Erfolg der im Lernprozess oder in den Lehrplänen eingeführten Änderungen zu bewerten. Sie muss in einen geplanten Evaluierungszyklus integriert werden, um nachzuweisen, dass die Intervention die beabsichtigten Ziele erreicht hat [10]. Nach Wilkes und Bligh [10] sollte die Bewertung den Erwerb von Fähigkeiten, Kenntnissen und Einstellungen umfassen, die auf die vorgeschlagenen Lernergebnisse ausgerichtet sind, um den Wandel des Lehrplans voranzutreiben und zu unterstützen.

Im Think4Jobs Projekt wurde beschlossen, einen studierendenorientierten Ansatz zu verwenden, der auf die Verbesserung der CT-Fähigkeiten und CT-Dispositionen innerhalb eines kursbezogenen Kontexts abzielt. Die Wirksamkeit der Lerninterventionen zur Förderung der CT-Kompetenzen wurde mit Hilfe von Fragebögen evaluiert, die zu diesem Zweck entwickelt wurden und mit den CT-bezogenen Lernzielen, die für jeden Kurs vorgeschlagen wurden, in einem kurzfristigen Bewertungsansatz abgestimmt waren.

In der Literatur gibt es mehrere formale standardisierte CT-Tests (wie den CCTT - Cornell Critical Thinking Test; das California Critical Thinking Dispositions Inventory - CCTDI; oder den Halpern Critical Thinking Assessment Test - HCTA, unter anderem) [11], die zur Bewertung von CT verwendet werden könnten, auch wenn die Instrumente häufiger die Bewertung von Fähigkeiten als von Dispositionen betreffen. Die konstruktive Übertragbarkeit vieler standardisierter CT-Tests auf verschiedene Bevölkerungsgruppen oder Disziplinen wurde jedoch in Frage gestellt [12, 13]. Zusammen mit den vielfältigen Konzeptualisierungen von CT könnte dies erklären, warum eine Einigung über ein Standardinstrument zur konsistenten

Bewertung der Lerneffektivität noch aussteht. Außerdem werden ständig neue Instrumente entwickelt und in verschiedenen Disziplinen und kulturellen Kontexten getestet.

Weitere Nachteile werden mit den standardisierten CT-Tests in Verbindung gebracht: Sie sind nicht überall leicht zugänglich; einige erfordern die Auswertung und Bewertung durch Experten und die Schulung von Bewertern, um individuelle Verzerrungen zu minimieren [14]. Einige von ihnen sind zu teuer, um routinemäßig eingesetzt zu werden [15], während andere Situationen darstellen, die nicht authentisch und weit von der Realität der Studierenden entfernt sind [11, 12]. Außerdem konzentrieren sich die Standardtests in der Regel besonders auf die Fähigkeiten oder erfassen Fähigkeiten und Neigungen in Kombination, und sie sind in der Regel lang und benötigen zwischen 50 und 80 Minuten zum Ausfüllen [12], was zu einer geringen Motivation der Studierenden führt, die Fragebögen auszufüllen [16]. Darüber hinaus ist bei einigen Tests unklar, wie das Denken der Befragten diskretere Dispositionen wie Aufgeschlossenheit oder Wissbegierde identifizieren kann [11] oder wie sie zwischen der Anwendung disziplinspezifischer Argumentationsfähigkeiten unterscheiden können, die von Studierenden in fortgeschritteneren Jahren ihres akademischen Weges erworben wurden.

Selbsteinschätzungsfragebögen zur Bewertung der Wahrnehmung des kritischen Denkens (CT) durch die Studierenden wurden in den letzten Jahrzehnten für verschiedene Disziplinen entwickelt und validiert, auch wenn ihre Verwendung nach wie vor nicht konsensfähig ist [17, 18], insbesondere aufgrund der geringen Zuverlässigkeit der Konstrukte, wenn sie in verschiedenen Populationen oder Disziplinen verwendet werden. Trotz der Kontroverse werden Selbstauskunftsfragebögen häufig verwendet, um Veränderungen im CT nach der Einführung neuer Lehrmethoden zu bewerten. In diesem Zusammenhang können sie Vorteile bei der Bewertung nicht-kognitiver Kompetenzen bieten, insbesondere wenn sie sich nicht direkt auf die Noten der Studierenden auswirken, sondern lediglich dazu dienen, die Leistungen der Studierenden zu überwachen oder zu verbessern und den individuellen Schulungsbedarf zu ermitteln [19].

Was die Bewertung von CT-Dispositionen betrifft, so gibt es nur wenige Instrumente [20, 21], und die wenigen verfügbaren basieren auf unterschiedlichen Konzeptualisierungen der

Einstellungsdimensionen, je nach Kontext der Studien. Wie bei den Instrumenten zur Messung der CT-Fähigkeiten mangelt es einigen der Skalen an Konsistenz [20].

Daher entschied sich das Think4Jobs-Konsortium für einen Selbstauskunftsfragebogen, da dieser leicht zu handhaben ist, wiederholt in einem Kurs eingesetzt werden kann und von Studierenden aus verschiedenen Fachrichtungen und Ländern gerne verwendet wird. Nachdem der Facione-Framework verwendet wurde, um die CT-Fähigkeiten und -Dispositionen zu identifizieren, auf die die Bildungsinterventionen abzielen, wie in IO3 [1] dargelegt, beschränkte das Konsortium die Suche auf Fragebögen, die vorzugsweise auf die Bewertung der CT-Fähigkeiten und -Dispositionen abzielen, wie sie im Facione-Framework konzeptualisiert sind [22, 23].

## 2. Methodologie

In Anbetracht der Tatsache, dass das Projekt beabsichtigt, Fähigkeiten und Dispositionen im Zusammenhang mit dem kritischen Denken (CT) zu messen, und weil Fähigkeiten und Dispositionen selten zusammen evaluiert werden, wurde beschlossen, zwei verschiedene Instrumente zu entwickeln (eines für die CT-Fähigkeiten und eines für die Dispositionen), die in einem einzigen Instrument zusammengefasst werden sollten.

Auf der Grundlage der bisherigen Erfahrungen der Partner und der verfügbaren Literatur wurden einige Hauptkriterien für die Auswahl des Instruments festgelegt: Es sollte ein geschlossener Test sein; er sollte einfach online zu verwalten sein; er sollte den vorgeschlagenen CT-Fähigkeiten und -Dispositionen entsprechen, die als Ergebnisse für die Aktivitäten in den Pilotkursen identifiziert wurden; er sollte für die Studierenden praktikabel sein; und er sollte keine hohen Anforderungen an das technische Fachwissen stellen, das zur Beantwortung und zum Abrufen von Informationen erforderlich ist. Darüber hinaus sollte der zusammengefasste Fragebogen ein zusätzliches Kriterium erfüllen, nämlich eine Begrenzung der Bearbeitungszeit auf 60 Minuten, um die Bereitschaft der Studierenden zur Nutzung des Instruments nicht zu beeinträchtigen.

## CTSAS-SF, das Instrument zur Bewertung der CT-Fähigkeiten im Think4Jobs-Projekt

Unter den verfügbaren Instrumenten zur Bewertung der Fähigkeiten zum kritischen Denken (CT) nach dem Konzept von Facione, wählte das Konsortium die von Nair [24] für Hochschulstuiierenden entwickelte Critical Thinking Self-Assessment Scale (CTSAS) aus. Der Fragebogen wurde in verschiedenen geografischen und kulturellen Kontexten getestet (was als eine Stärke des Instruments angesehen wurde) und erzielte gute Ergebnisse bei den Tests zur Zuverlässigkeit und internen Konsistenz sowie bei der konfirmatorischen Faktorenanalyse [25]. Der ursprüngliche CTSAS-Fragebogen bestand aus 115 Items, die auf einer siebenstufigen Ratingskala (von 0=nie bis 6=immer) bewertet wurden, und wurde als zu lang empfunden (der Autor berichtete von ca. 50 Minuten), vor allem weil geplant war, ihn mit einem anderen Instrument zur Erfassung von CT-Dispositionen zu verbinden. Daher wurde beschlossen, eine Kurzform des CTSAS-Fragebogens zu verwenden, die speziell für das Projekt entwickelt wurde.

Die ursprüngliche Nair-Skala wurde unter der Leitung von zwei portugiesischen Forschern in einem zweistufigen Verfahren gekürzt, wobei die folgenden Kriterien für die Ablehnung von Items angewandt wurden: 1. Die Items mit Ladungsgewichten unter 0,500 wurden eliminiert, so dass 84 Items übrigblieben. 2. Redundante Items und solche mit einem nicht-kognitiven Schwerpunkt wurden zur Eliminierung markiert, so dass 58 Items übrig blieben. Nach der Markierung der zu eliminierenden Items wurde das verbleibende Konstrukt von zwei unabhängigen Experten analysiert, um den ablehnenden Vorschlag auf der Grundlage der Konzeptualisierung der CT-Fähigkeiten und Teilfähigkeiten von Facione zu bestätigen oder zu verwerfen. Diese Experten empfahlen die Beibehaltung der Items 16 und 19 der ursprünglichen Skala aufgrund ihrer theoretischen Relevanz. Am Ende des Prozesses enthielt die CTSAS-Kurzform insgesamt 60 von Experten begutachtete Items. An den Items der ursprünglichen CTSAS-Skala wurden keine Änderungen vorgenommen (für Einzelheiten siehe [2]).

In der CTSAS-Kurzform (CTSAS-SF) schwankte die Anzahl der Items zur Bewertung jeder Dimension zwischen 7 und 13. Bei den Subdimensionen (oder Subskills) variierte die Anzahl der Items zwischen 3 und 7, mit Ausnahme von 5 Subdimensionen (*Bedeutung entschlüsseln, Argumente erkennen, Behauptungen bewerten, Ergebnisse angeben und Verfahren*

begründen), die nur zwei Items enthielten (Tabelle 2). Der CTSAS-SF behielt den Rahmen der ursprünglichen Skala bei, bei der die Studierenden mit der Frage "Was tun Sie, wenn Sie mit einem Problem konfrontiert werden?" beginnen und aufgefordert werden, die Items anhand einer siebenstufigen Likert-Skala mit folgender Entsprechung zu beantworten: 0= Nie; 1 = Selten; 2 = Gelegentlich; 3 = Gewöhnlich; 4 = Oft; 5 = Häufig; 6 = Immer.

**Tabelle 2 – Die Struktur und die Dimensionen des CTSAS-SF-Fragebogens**

CTSAS Dimensionen (Skills / Subskills)		Items in der CTSAS Kurzform
	Categorisierung	1 – 3
Interpretation	Bedeutung klären	6 – 9
	Bedeutung entschlüsseln	4, 5
Analyse	Erkennen von Argumenten	15, 16
	Analysieren von Argumenten	17 – 20
	Prüfung von Ideen	10 – 14
Bewertung	Bewertung von Behauptungen	21, 22
	Bewertung von Argumenten	23 – 27
Schlussfolgerung	Ziehen von Schlussfolgerungen	36 – 40
	Vermutungen über Alternativen anstellen	31 – 35
	Beweise abfragen	28 – 30
Erläuterung	Angabe der Ergebnisse	41, 42
	Verfahren rechtfertigen	43, 44
	Argumente vorbringen	45 – 50
Selbstregulierung	Selbstuntersuchung	51 – 57
	Selbstkorrektur	58 – 60

## SENCTDS, das Instrument zur Bewertung der CT-Dispositionen im Think4Jobs-Projekt

Wie bereits erwähnt, gibt es weitaus weniger Instrumente zur Messung von CT-Dispositionen als zur Bewertung von CT-Fähigkeiten. Unter den verfügbaren Fragebögen zur Bewertung von CT-Dispositionen hat das Konsortium die von Quinn et al. [3] entwickelte Student-Educator Negotiated Critical Thinking Dispositions Scale (SENCTDS) übernommen, die an einer gemischten irischen und amerikanischen Studentenpopulation validiert wurde. Die Skala wurde unter Berücksichtigung verschiedener CT-Dispositionen entwickelt, die die Autoren als wichtig für den Arbeitsmarkt und für die Entscheidungsfindung in der realen Welt erachteten [3]. Einige Items der Skala kombinieren einige der klassischen CT-Dispositionen von Facione zu neuen Dimensionen, von denen angenommen wird, dass sie für den akademischen Erfolg und den Erfolg auf dem Arbeitsmarkt wichtig sind (z. B. die Dispositionen *Organisation, Ausdauer und intrinsische Zielmotivation*). Die Items repräsentieren sechs Dispositionsdimensionen (*Reflexion, Aufmerksamkeit, Aufgeschlossenheit, Organisation, Durchhaltevermögen und intrinsische Zielmotivation*) und sind als Aussagen formuliert, zu denen sich die Studierenden auf einer 7-stufigen Likert-Skala positionieren müssen: 1 = stimme überhaupt nicht zu; 2 = stimme nicht zu; 3 = stimme kaum zu; 4 = stimme weder zu noch stimme ich zu; 5 = stimme kaum zu; 6 = stimme zu; 7 = stimme voll und ganz zu. Für die Skala spricht, dass sie gut mit anderen validierten Konstrukten korreliert, die zur Bewertung von CT-Dispositionen dienen [3].

Die SENCTDS wurde in ihrer ursprünglichen Version verwendet, die insgesamt 21 Items umfasst (Tabelle 3). Trotz des leichten Unterschieds bei der Bestimmung der Skalenpunkte im Vergleich zum CTSAS-SF wurde beschlossen, die ursprüngliche Darstellung der Likert-Skala beizubehalten.

## Übersetzung der Instrumente zur Bewertung der CT-Fähigkeiten und CT-Dispositionen

Die angenommenen CTSAS\_SF und SENCTDS, die ursprünglich in englischer Sprache verfasst waren, wurden ins Portugiesische, Rumänische, Griechische und Deutsche übersetzt. Die

Übersetzung in diese Sprachen erfolgte nach den empfohlenen Verfahren (Übersetzung, Überarbeitung und Verfeinerung), um sicherzustellen, dass die Bedeutung, Konnotation und Konzeptualisierung dem Originalinstrument entsprachen [26, 27]. Zwei zweisprachige Übersetzer aus jedem Land, das eine nicht-englische Version des Fragebogens verwendete, übersetzten die angenommenen Instrumente in ihre Muttersprache; verschiedene Forschergruppen analysierten dann die Übersetzungen, um Unterschiede zwischen den beiden Versionen des Fragebogens zu ermitteln und die Genauigkeit der Übersetzung und ihre Übereinstimmung mit dem Original sicherzustellen [28]. Die übersetzten Versionen wurden dann von einer Gruppe von Experten aus jedem nationalen Team des Projekts überprüft, die die inhaltliche Gleichwertigkeit des Instruments beurteilten. Die Konkordanz der Experten wurde als gleichwertige Bewertung des übersetzten Fragebogens angesehen.

**Tabelle 3** – Die Struktur und die Dimensionen des SENCTDS-Fragebogens

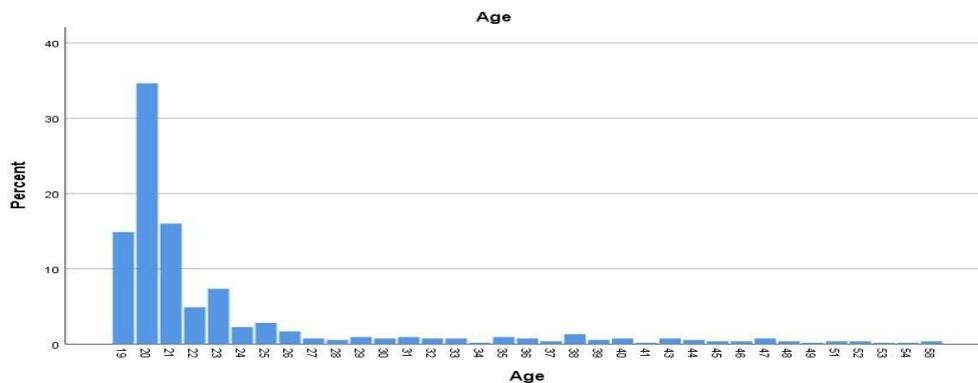
SENCTDS Dimensionen (Dispositionen)	Items in SENCTDS
Reflektion	1 – 3
Achtsamkeit	4 – 7 (rückwärts gepunktet)
Aufgeschlossenheit	8 – 11 (rückwärts gepunktet)
Organisation	12 – 14
Durchhaltevermögen	15 – 17
Intrinsische Motivation	18 – 21

### Datenerhebung und Analyse

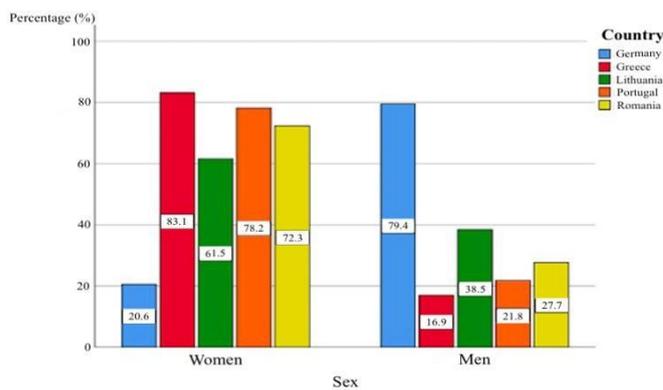
Für die Validierung der Fragebögen wurden alle ersten Antworten der Teilnehmer an den gemischten Kursen verwendet, die entweder während des Herbst-/Wintersemesters oder des Frühjahrs-/Sommersemesters in den fünf am Projekt beteiligten Hochschuleinrichtungen durchgeführt wurden.

Die Antworten von 531 Universitätsstudenten (389 weiblich, 142 männlich) im Alter von 19 bis 58 Jahren (Mittelwert = 23,47; SD = 7,184) wurden in dieser Analyse berücksichtigt.

Abbildung 1 zeigt die Altersverteilung in der Population der Befragten, und Abbildung 2 zeigt die unausgewogene Geschlechterverteilung pro Land.



**Abbildung 1** – Altersverteilung in der befragten Bevölkerung (n=531)



**Abbildung 2** – Geschlechterverteilung nach Ländern (n=531).

Es wurde eine ungleichmäßige Verteilung der Teilnehmer nach Ländern festgestellt: 33,3% kamen aus Griechenland, 29,4% aus Portugal, 21,1% aus Rumänien, 9,8% aus Litauen und 6,4% aus Deutschland. Die Studierenden beantworteten die Fragebögen in Griechisch, Portugiesisch, Rumänisch, Englisch bzw. Deutsch. Die Studierenden besuchten die folgenden Studiengänge (Disziplinen): Wirtschaftsinformatik (Deutschland), Lehramt (Griechenland), Veterinärmedizin (Portugal) und Betriebs- und Volkswirtschaft (Rumänien) sowie den Studiengang Englisch als Fremdsprache (Litauen).

Die Studierenden unterzeichneten eine Einverständniserklärung, die dem Fragebogen beigefügt war, und wurden darüber informiert, dass sie jederzeit aus der Studie aussteigen

konnten, ohne dass ihnen eine Strafe oder ein Leistungsverlust drohte. Bei der Studienpopulation handelte es sich um eine nicht-randomisierte Zufallsstichprobe, die sich aus den freiwilligen Antworten der in den von Think4Jobs konzipierten CTBACs eingeschriebenen Studierenden ergab.

Die CTSAS-SF- und SENCTDS-Instrumente wurden in einem einzigen Formular zusammengefasst, das auf der Google Forms-Plattform zur Verfügung gestellt wurde, nachdem eine Einladung über die Moodle-Seite des Kurses, für den sie eingeschrieben waren, verschickt wurde. Der Prozess wurde von den an den Pilotkursen beteiligten Lehrkräften überwacht. Das Formular enthielt auch einen vorläufigen Abschnitt zur allgemeinen Identifizierung der Studierenden (E-Mail, Name, Land, Fachrichtung und Kurs, Geschlecht und Alter). Die Antworten wurden von der Google-Plattform in eine Excel-Datei *pro* Land übertragen. Die E-Mail und der Name - die notwendig sind, um aufeinanderfolgende Antworten zeitlich zuzuordnen - wurden für die Analyse der Ergebnisse während des Anonymisierungsschritts der Datenbankvorbereitung entfernt. Die Namen wurden in einen alphanumerischen Code umgewandelt (bestehend aus dem Code für das Land - GR, LT, RO, GE und PT für Griechenland, Litauen, Rumänien; Deutschland und Portugal - plus einer fortlaufenden Nummer von 1 bis n), und die Spalte mit den Namen und E-Mails wurde gelöscht. Die Anonymisierung der Daten wurde von einem anderen Forscher als demjenigen vorgenommen, der die statistische Analyse durchführte, um das Risiko einer Verzerrung zu verringern. Jede Länderdatenbank wurde auf inkonsistente Daten überprüft, bevor sie für die statistische Analyse zu einer einzigen Datenbank zusammengeführt wurden.

Die statistische Analyse umfasste die deskriptiven Maße der Items (Mittelwert, Standardabweichung, Schiefe, Kurtosis), den Kolmogorov-Smirnov-Test auf Gleichverteilung und Mann-Witney's U für mittlere Rangunterschiede. Um zu beurteilen, ob der CTSAS-SF und SENCTDS dem ursprünglichen Faktorenmodell entsprechen, wurde eine konfirmatorische Faktorenanalyse (CFA) mit gewichteten kleinsten quadratischen Mittelwerten und Varianzen (WLSMV) als Schätzmethode aufgrund der ordinalen Natur der Daten durchgeführt [29]. Zu den durchgeführten Modellanpassungsindizes gehören der  $\chi^2$ -Test auf exakte Anpassung, der Comparative Fit Index (CFI), der Tucker-Lewis Index (TLI) und der Root Mean Squared Error of

Approximation (RMSEA). In Anlehnung an Hu und Bentler [30] betrachten wir CFI- und TLI-Werte  $\geq .90$  und RMSEA  $\leq .06$  (90%IC) als akzeptable Anpassungswerte. Die Daten wurden in dem Modell als ordinal spezifiziert.

Die Reliabilität und interne Konsistenz der Skala und der Subskalen wurden anhand von Cronbachs Alpha geschätzt. Nach Hair et al. [31] wurden Alphas über .70 als gute Reliabilitätsindizes angesehen. Die Multigruppeninvarianz wurde für Geschlecht und Länder/Disziplinen bewertet. Unterschiede zwischen den RMSEA- und CFI-Werten unter .015 bzw. .01 wurden als Kriterium für Invarianz verwendet [32, 33]. Die univariate deskriptive und interne Konsistenz wurde mit IBM SPSS Statistics 26 berechnet. Die CFA und die Multigruppen-Invarianzanalyse wurden mit MPlus 7.4 [34] durchgeführt.

### 3. Ergebnisse - Validierung der Skalen für CT-Fähigkeiten und CT-Dispositionen

Die Ergebnisse werden in drei Abschnitten sowohl für die Skalen der CT-Fähigkeiten als auch der CT-Dispositionen (CTSAS-SF bzw. SENCTDS) dargestellt. Der erste Abschnitt enthält deskriptive Statistiken zu den Items. Der zweite Abschnitt zeigt die Ergebnisse der konfirmatorischen Faktorenanalyse. Der dritte Abschnitt zeigt die Analyse der Mehrgruppeninvarianz.

#### Deskriptive Analyse der Items in den Skalen CTSAS-SF und SENCTDS

Beim CTSAS-SF liegt der Mittelwert der 60 Items zwischen 3,13 ("*Ich schreibe Aufsätze mit angemessenen Argumenten und Begründungen für eine bestimmte Politik oder Situation*") und 5,04 ("*Ich versuche, den Inhalt des Problems zu verstehen*"). Die Standardabweichung schwankt zwischen 0,958 ("*Ich versuche, den Inhalt des Problems herauszufinden*") und 1,734 ("*Ich schreibe Aufsätze mit angemessenen Argumenten, die mit Gründen für eine bestimmte Politik oder Situation untermauert sind*"). Der K-S-Test zeigt, dass die Daten gleichmäßig auf weibliche und männliche Studierenden verteilt sind ( $p > .050$ ), mit Ausnahme des Items "*Ich kann Ergebnisse logisch darstellen, um ein bestimmtes Problem zu lösen*" ( $Z = 1.533$ ;  $p = .018$ )

und des Items *"Ich reagiere auf begründete Kritik, die man gegen die eigenen Standpunkte vorbringen könnte"* ( $Z = 1.772$ ;  $p = .004$ ). Die Beschreibung der Items ist in der ergänzenden Tabelle 1 dargestellt.

Der Mann-Witney's U-Test zeigt keine statistisch signifikanten Unterschiede in den CT-Fähigkeiten (CTSAS-SF) zwischen weiblichen und männlichen Studierenden ( $p > .050$ ) mit Ausnahme der Items *"Ich beobachte den Gesichtsausdruck, den Menschen in einer bestimmten Situation verwenden"* (Std U =  $-2.230$ ;  $p = .026$ ), *"Ich kann Ergebnisse logisch darstellen, um ein gegebenes Problem zu lösen"* (Std U =  $2.382$ ;  $p = .017$ ), *"Ich reagiere auf begründete Kritik, die man gegen den eigenen Standpunkt vorbringen könnte"* (Std U =  $3.957$ ;  $p < .001$ ) und *"Ich begründe die Zurückweisung der Behauptung eines anderen"* (Std U =  $2.588$ ;  $p = .010$ ).

Bei der SENCTDS-Skala schwankt der Mittelwert der 21 Items zwischen  $3,83$  (*"Ich finde, dass ich leicht abgelenkt bin, wenn ich über eine Aufgabe nachdenke"*) und  $5,85$  (*"Wenn ich vor einer Entscheidung stehe, suche ich so viele Informationen wie möglich"*). Die Standardabweichung schwankt zwischen  $1,070$  (*"Wenn mir eine Theorie, Interpretation oder Schlussfolgerung präsentiert wird, versuche ich zu entscheiden, ob es gute Belege dafür gibt"*) und  $1,827$  (*"Es fällt mir schwer, mich beim Nachdenken über Probleme zu konzentrieren"*). Der K-S-Test zeigt, dass die Daten bei weiblichen und männlichen Studenten gleich verteilt sind ( $p > .050$ ), mit Ausnahme der Items *"Ich verpasse oft wichtige Informationen, weil ich an andere Dinge denke"* ( $Z = 1.370$ ;  $p = .047$ ), *"Ich mache gerne Listen mit Dingen, die ich tun muss und Gedanken, die ich habe"* ( $Z = 1.920$ ;  $p = .001$ ), *"Ich mache mir Notizen, um meine Gedanken zu ordnen"* ( $Z = 1.891$ ;  $p = .002$ ), *"Ich mache einfache Diagramme oder Tabellen, um große Mengen an Informationen zu ordnen"* ( $Z = 1.598$ ;  $p = .012$ ) und *"Ich freue mich darauf, herausfordernde Dinge zu lernen"* ( $Z = 1.436$ ;  $p = .032$ ). Die Beschreibung der Items findet sich in der ergänzenden Tabelle 2.

Der Mann-Witney's U-Test zeigt keine statistisch signifikanten Unterschiede in den CT-Dispositionen (SENCTDS) zwischen weiblichen und männlichen Studierenden ( $p > 0,05$ ), außer bei neun Items: *"Wenn ich vor einer Entscheidung stehe, suche ich so viele Informationen wie möglich"* (Std U =  $-1.104$ ;  $p = .028$ ), *"Es fällt mir schwer, mich zu konzentrieren, wenn ich über*

*Probleme nachdenke*" (Std U = 2.819;  $p = .005$ ), "*Ich verpasse oft wichtige Informationen, weil ich an andere Dinge denke*" (Std U = 2.426,  $p = .015$ ), "*Ich mache mir gerne Listen mit Dingen, die ich tun muss und Gedanken, die ich habe*" (Std U = -4.577;  $p \leq .0001$ ), "*Ich mache mir Notizen, um meine Gedanken zu ordnen*" (Std U = -5,010;  $p < .0001$ ), "*Ich mache einfache Diagramme oder Tabellen, um große Mengen an Informationen zu ordnen*" (Std U = -3,557;  $p \leq .0001$ ), "*Ich genieße Informationen, die mich zum Nachdenken herausfordern*" (Std U = 1,964;  $p = .050$ ), "*Ich freue mich darauf, herausfordernde Dinge zu lernen*" (Std U = 2,804;  $p = .005$ ), und "*Die Erledigung schwieriger Aufgaben macht mir Spaß*" (Std U = 2,515,  $p = .012$ ).

### **Konfirmatorische Faktorenanalyse (CFA) und Zuverlässigkeit**

Mit der CFA soll bestätigt werden, dass die Fragebögen dem ursprünglichen Modell entsprechen, das für die ursprünglichen Skalen vorgeschlagen wurde. Für eine umfassende Analyse der Struktur und der Beziehungen der Items in jedem Instrument wurden sechs latente Fähigkeiten und ein allgemeines Konstrukt anhand von fünf aufeinanderfolgenden Modellen mit zunehmender Komplexität getestet.

- Modell 1: Ein-Faktor-Modell. Dieses Modell testet das Vorhandensein eines globalen Faktors für CT-Fähigkeiten oder Dispositionen, der die Varianzen der Items erklärt;
- Modell 2: Sechs-Faktoren-Modell (nicht korreliert). Dieses Modell testet die Existenz von 6 nicht korrelierten Faktoren, die die Varianz der Items in jeder Skala erklären;
- Modell 3: Sechs-Faktoren-Modell (korreliert). Dieses Modell testet die Existenz von 6 korrelierten latenten Faktoren, von denen jeder die Varianz einer Gruppe von Items erklärt;
- Modell 4: Faktorenmodell zweiter Ordnung. Dieses Modell testet, wie das globale Konstrukt der kritischen Denkfähigkeiten oder Dispositionen die Varianz der 6 latenten Fähigkeiten erklärt, die ihrerseits jeweils eine Reihe von Items erklären;
- Modell 5: Zwei-Faktoren-Modell. Dieses Modell testet die Möglichkeit, dass die Varianzen der Skalenitems durch ein globales Konstrukt für kritische Denkfähigkeiten oder Dispositionen und durch die 6 latenten Fähigkeiten unabhängig voneinander erklärt werden.

Tabelle 4 zeigt die Modellanpassungsindizes für jeden Fragebogen. Im Fall der CTSAS-SF-Skala sind die Anpassungsindizes für die Modelle 3 und 4 zufriedenstellend, nicht aber für die Modelle 1, 2 und 5. Da die Modelle 3 und 4 nicht verschachtelt sind, stützen wir unsere Interpretation auf die Unterschiede in den Anpassungsindizes. Der Differenzwert der RMSEA- und CFI-Indizes zwischen Modell 3 (das die besten Anpassungsindizes aufweist) und Modell 4 (das das von Nair [24] vorgeschlagene Originalmodell darstellt) liegt unter .015 bzw. .010 ( $\Delta$ RMSEA = .002;  $\Delta$ CFI = .003), was darauf hindeutet, dass beide Modelle zur Validierung der internen Struktur des Fragebogens verwendet werden können. Da Modell 4 das ursprüngliche Modell darstellt, wird es als angepasste Faktorenstruktur akzeptiert und für die folgende Analyse berücksichtigt.

Die Faktorenladungen für die CTSAS-SF-Skala sind in der zusätzlichen Tabelle 3 dargestellt. Die Ladungen sind signifikant ( $p < .001$ ) und variieren von 0,386 ("Ich beobachte den Gesichtsausdruck, den Menschen in einer bestimmten Situation verwenden") bis 0,786 ("Ich überprüfe und überdenke ständig Strategien, um mein Denken zu verbessern"). Alle Faktorladungen liegen über 0,500, mit Ausnahme der Items "Ich beobachte den Gesichtsausdruck, den Menschen in einer bestimmten Situation machen" (0,386), "Ich kläre meine Gedanken, indem ich sie jemand anderem erkläre" (0,422) und "Ich lehne eine alternative Lösung selbstbewusst ab, wenn sie keine Beweise enthält" (0,470).

Im Fall der SENCTDS-Skala war das einzige Modell mit einer zufriedenstellenden Anpassungsgüte Modell 3, das zur Validierung der internen Struktur des Fragebogens verwendet wurde. Für diese Skala sind die Faktorladungen in der ergänzenden Tabelle 4 dargestellt. Alle Ladungen sind signifikant und variieren zwischen 0,659 ("*Beim Denken geht es nicht darum, 'flexibel' zu sein, sondern darum, 'richtig' zu sein*") und 0,908 ("*Ich mache mir Notizen, damit ich meine Gedanken ordnen kann*"). Alle Faktorladungen liegen über 0,600.

Das CTSAS-SF-Instrument weist eine ausgezeichnete interne Konsistenz auf (Cronbachs  $\alpha = 0,969$ ). Die Cronbachs Alphas für jede Dimension der Skala liegen über 0,700, was auf eine gute faktorielle Reliabilität hinweist (Tabelle 4). Die Korrelationen zwischen den Faktoren und zwischen den Faktoren sind stark (von 0,750 bis 0,965) (Tabelle 5). Alle Korrelationen sind bei einem p-Wert  $\leq .0001$  signifikant.

Die SENCTDS-Skala weist eine sehr gute interne Konsistenz auf (Cronbachs  $\alpha = 0,842$ ). Die Cronbachs Alphas für jede Dimension der Skala liegen über .700, was eine gute faktorielle Reliabilität zeigt (Tabelle 4). Die Korrelationen zwischen den Faktoren variieren jedoch von mittelmäßig bis stark (von .135 bis .769), mit Ausnahme der nicht signifikanten Korrelationen zwischen Organisation und Aufmerksamkeit oder Aufgeschlossenheit (Tabelle 5). Alle Korrelationen sind bei einem p-Wert < .010 signifikant.

**Tabelle 4.** Goodness-of-fit-Indizes für die Instrumente CTSAS-SF und SENCTDS.

Skala	Modell	$\chi^2$ (df)	$p$	RMSEA [90%IC]	CFI	TLI
CTSAS-SF	Modell 1: Ein-Faktor Modell	5159.412 (1710)	<.0001	.061 [.059-.063]	.893	.890
	Modell 2: 6-Faktoren Modell (nicht korreliert)	29275.338 (1710)	<.0001	.174 [.172-.176]	.148	.118
	Modell 3: 6-Faktoren Modell (korreliert)	3871.243 (1695)	<.0001	.049 [.047-.051]	.933	.930
	Modell 4: Faktorenmodell zweiter Ordnung	3975.885 (1704)	<.0001	.051 [.049-.053]	.927	.924
	Modell 5: Zwei-Faktoren Modell	18656.904 (1657)	<.0001	0.139 [.137-.141]	.474	.439
SENCTDS	Modell 1: Ein-Faktor Modell	4655.783 (189)	<.0001	.211 [.206-.216]	.579	.532
	Modell 2: 6-Faktoren Modell (nicht korreliert)	3828.759 (189)	<.0001	.190 [.185-.196]	.657	.618
	Modell 3: 6-Faktoren Modell (korreliert)	447.677 (174)	<.0001	.054 [.048-.061]	.974	.969
	Modell 4: Faktorenmodell zweiter Ordnung	686.865 (183)	<.0001	.072 [.066-.078]	.952	.945
	Modell 5: Zwei-Faktoren Modell	676.807 (165)	<.0001	.076 [.070-.082]	.952	.939



**Tabelle 5.** Cronbach's Alpha Reliabilitätsindex für beide Skalen und Korrelationen zwischen den Faktoren plus den Faktoren und dem allgemeinen CT-Fähigkeitskonstrukt für die CTSAS-SF-Skala (nach dem Faktorenmodell zweiter Ordnung) und Korrelationen zwischen den Faktoren für die SENCTDS-Skala (nach dem Modell mit sechs korrelierten Faktoren).

<b>Fähigkeiten</b>	<b><math>\alpha</math></b>	<b>CT-Fähigkeits</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
1. Interpretation	.772	.881					
2. Analyse	.888	.925	.905				
3. Bewertung	.858	.965	.810	.934			
4. Schlussfolgerung	.905	.956	.806	.858	.937		
5. Erläuterung	.853	.907	.765	.825	.864	.868	
6. Selbstregulierung	.905	.851	.750	.750	.781	.841	.805

<b>Dispositionen</b>	<b><math>\alpha</math></b>		<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
1. Reflektion	.796	--					
2. Achtsamkeit	.853	--	.135				
3. Aufgeschlossenheit	.773	--	.251	.396			
4. Organisation	.772	--	.431	-.020 (ns)	.077 (ns)		
5. Durchhaltevermögen	.792	--	.621	.319	.256	.355	
6. Intrinsische Motivation	.842	--	.567	.265	.321	.315	.769

## Mehrgruppeninvarianz für Geschlecht

Um die faktorielle Strukturinvarianz beider Fragebögen über die Geschlechter hinweg zu überprüfen, wurde eine Multigruppeninvarianzanalyse unter Berücksichtigung des Faktorenmodells zweiter Ordnung im Fall des CTSAS-SF und des Modells mit sechs korrelierten Faktoren im Fall des SENCTDS durchgeführt. Aufgrund der ordinalen Natur der Daten wurde die WLSMV als Schätzmethode verwendet. Der erste Schritt des Verfahrens bestand darin, eine Baseline für beide Gruppen (weibliche und männliche Studenten) zu erstellen, wobei für jede Gruppe unabhängige CFAs verwendet wurden. Anschließend wurde eine CFA für beide Gruppen gleichzeitig durchgeführt, um die Invarianz zu testen. Die drei Invarianzmodelle, vom weniger restriktiven (das konfigurative Modell) bis zum restriktivsten (die skalare Invarianz), wurden getestet. Die Ergebnisse sind in Tabelle 6 dargestellt.

**Tabelle 6.** Die Goodness-of-fit-Indizes für die Mehrgruppeninvarianz nach Geschlecht für den CTSAS-SF (Faktorenmodell zweiter Ordnung) und den SENCTDS (Modell mit sechs korrelierten Faktoren).

<b>CTSAS-SF</b>	<b>Baseline Modelle</b>	<b><math>\chi^2</math> (df)</b>	<b>p</b>	<b>RMSEA [90%IC]</b>	<b>CFI</b>	<b>TLI</b>
	Weiblich	3488.157 (1704)	<.0001	.052 [.049-.054]	.929	.926
	Männlich	2314.349 (1704)	<.0001	.050 [.045-.055]	.948	.946
	<b>Invarianz</b>	<b><math>\chi^2</math> (df)</b>	<b>p</b>	<b>RMSEA [90%IC]</b>	<b>CFI</b>	<b>TLI</b>
	Konfigurative Invarianz	5521.460 (3390)	<.0001	.049 [.046-.051]	.939	.936
	Metrische Invarianz	5490.717 (3444)	<.0001	.047 [.045-.050]	.941	.940
	Skalare Invarianz	5613.987 (3732)	<.0001	.044 [.041-.046]	.946	.949
	<b>Modelvergleich</b>	<b><math>\chi^2</math> (df)</b>	<b>p</b>	<b><math>\Delta</math>RMSEA</b>	<b><math>\Delta</math>CFI</b>	
	Metrisch vs. Konfigurativ	45.988 (54)	.773	.002	.002	
	Skalar vs. Konfigurativ	370.658 (342)	.137	.005	.007	
Skalar vs. Metrisch	328.786 (288)	.049	.003	.005		
<b>SENCTDS</b>	<b>Baseline Modelle</b>	<b><math>\chi^2</math> (df)</b>	<b>p</b>	<b>RMSEA [90%IC]</b>	<b>CFI</b>	<b>TLI</b>
	Weiblich	352.859 (174)	<.0001	.051 [.044-.059]	.977	.973
	Männlich	313.264 (174)	<.0001	.075 [.062-.088]	.953	.943
	<b>Invarianz</b>	<b><math>\chi^2</math> (df)</b>	<b>p</b>	<b>RMSEA [90%IC]</b>	<b>CFI</b>	<b>TLI</b>
	Konfigurative Invarianz	660.692 (348)	<.0001	.058 [.051-.065]	.970	.964
	Metrische Invarianz	663.475 (363)	<.0001	.056 [.049-.063]	.971	.966
	Skalare Invarianz	754.103 (460)	<.0001	.049 [.043-.055]	.972	.974

	<i>Modellvergleich</i>	$\chi^2$ (df)	<i>p</i>	$\Delta$ RMSEA	$\Delta$ CFI
<b>SENCTDS</b>	<i>Metrisch vs. Konfigurativ</i>	14.683 (15)	.4745	.002	.001
	<i>Skalar vs. Konfigurativ</i>	123.360 (112)	.2180	.009	.002
	<i>Skalar vs. Metrisch</i>	110.600 (97)	.1632	.007	.001

Auf der Grundlage der Anpassungsgüte der verschiedenen getesteten Invarianzmodelle (konfigurative, metrisch und skalar) im CTSAS-SF-Instrument wird die Stabilität der Faktorenstruktur bei beiden Geschlechtern bestätigt. Die Differenz ( $\Delta$ ) der CFI- und RMSEA-Werte zwischen den Modellen beträgt weniger als 0,015 bzw. 0,010, was die Invarianz der Faktorenstruktur, die Invarianz der Faktorladungen und die Invarianz der Itemabschnitte beim Vergleich zwischen weiblichen und männlichen Studierenden belegt. In ähnlicher Weise zeigt das SENCTDS-Instrument eine gute Stabilität der Faktorenstruktur bei beiden Geschlechtern. Die Differenz der CFI- und RMSEA-Werte zwischen den Modellen beträgt weniger als 0,010 bzw. 0,015 [32, 33], was die Invarianz der Faktorenstruktur, der Faktorladungen und der Item-Intercepts beim Vergleich von weiblichen und männlichen Studierenden bestätigt. Nachdem die Instrumenteninvarianz bestätigt wurde, wurden die strukturelle Invarianz in Bezug auf die Heterogenität der Population sowie die latente Mittelwertinvarianz getestet. Strukturelle Invarianz prüft, ob das Kovarianzniveau zwischen den Faktoren für beide Gruppen gleich ist. Die latente Mittelwertinvarianz bewertet, ob die latenten Mittelwerte in beiden Gruppen gleich sind.

Tabelle 7 zeigt die Ergebnisse der strukturellen Invarianz in beiden Gruppen in CTSAS-SF und SENCTDS. Der Wald-Test zeigt einen signifikanten Unterschied zwischen den Faktorkorrelationen der weiblichen und männlichen Modelle in CTSAS-SF (Wald = 6,507; df = 1; p = 0,011), aber nicht in SENCTDS (Wald=0,316; df=1; p=0,5742). In der CTSAS-SF sind die Faktorkovarianzen im männlichen Modell signifikant höher als im weiblichen Modell, was auf eine gewisse Heterogenität der Bevölkerung schließen lässt. Im Gegensatz dazu sind bei SENCTDS die Faktorkovarianzen zwischen Männern und Frauen ähnlich, was auf ein einheitlicheres Verhalten zwischen den Geschlechtern hindeutet.

**Tabelle 7.** Faktorkovarianzen nach Geschlecht für den CTSAS-SF (CT-Fähigkeiten) und den SENCTDS (CT-Dispositionen).

<i>Fähigkeiten</i>	<i>Interpretation</i>		<i>Analyse</i>		<i>Bewertung</i>		<i>Schlussfolgerung</i>		<i>Erläuterung</i>	
	<i>F</i>	<i>M</i>	<i>F</i>	<i>M</i>	<i>F</i>	<i>M</i>	<i>F</i>	<i>M</i>	<i>F</i>	<i>M</i>
<i>Analyse</i>	.888	.941								
<i>Bewertung</i>	.760	.900	.922	.955						
<i>Schlussfolgerung</i>	.759	.890	.838	.902	.924	.956				
<i>Erläuterung</i>	.739	.849	.816	.877	.850	.907	.856	.925		
<i>Selbstregulierung</i>	.720	.808	.738	.780	.759	.825	.805	.907	.782	.885

<i>Dispositionen</i>	<i>Reflektion</i>		<i>Achtsamkeit</i>		<i>Aufgeschlossenheit</i>		<i>Organisation</i>		<i>Durchhaltevermögen</i>	
	<i>F</i>	<i>M</i>	<i>F</i>	<i>M</i>	<i>F</i>	<i>M</i>	<i>F</i>	<i>M</i>	<i>F</i>	<i>M</i>
<i>Achtsamkeit</i>	.119**	.220***								
<i>Aufgeschlossenheit</i>	.242	.286	.428	.348						
<i>Organisation</i>	.561	.528	-.003 (ns)	.017 (ns)	.151*	-.144 (ns)				
<i>Durchhaltevermögen</i>	.559	.553	.246	.495	.324	.109 (ns)	.401	.281		
<i>Intrinsische Motivation</i>	.572	.589	.232	.309	.387	.196***	.382	.338	.767	.807

F = weibliche Studenten, M = männliche Studenten. Alle Korrelationen sind auf dem p-Niveau < .001 für die CTSAS-SF-Skala signifikant.

Bei SENCTDS sind alle Korrelationen auf einem p-Niveau < .001 signifikant, außer für: \* p=.003; \*\*p=.024; \*\*\*=.005; ns= nicht-signifikant

Bei der Analyse der Mittelwertinvarianz bilden die weiblichen Studierenden die Ausgangsgruppe mit einem latenten Mittelwert von Null. Die Mittelwertvergleiche sowohl für den CTSAS-SF als auch für den SENCTDS sind in Tabelle 8 dargestellt. Im CTSAS-SF gibt es keine signifikanten Unterschiede in den Mittelwerten der Faktoren zwischen Frauen und Männern, aber im SENCTDS weisen Männer signifikant höhere Durchschnittswerte für *Achtsamkeit* und *Intrinsische Motivation* auf und einen signifikant niedrigeren Durchschnittswert für *Organisation*.

**Tabelle 8.** Latente Mittelwertunterschiede zwischen weiblichen und männlichen Studierenden.

<b>Fähigkeiten</b>		<b><math>\Delta</math>Means</b>	<b>SE</b>	<b>Est/SE</b>	<b>p</b>
CTSAS-SF	1. Interpretation	-.014	.106	-.129	.897
	2. Analyse	.023	.096	.244	.807
	3. Bewertung	.071	.096	.736	.462
	4. Schlussfolgerung	-.051	.099	-.512	.608
	5. Erläuterung	.177	.097	1.832	.067
	6. Selbstregulierung	-.005	.098	-.046	.963
<b>Dispositionen</b>					
SENCTDS	1. Reflektion	-.197	.118	-1.676	.094
	2. Achtsamkeit	.206	.103	1.994	<b>.046</b>
	3. Aufgeschlossenheit	-.120	.108	-1.111	.266
	4. Organisation	-.511	.110	-4.647	<b>≤.0001</b>
	5. Durchhaltevermögen	.024	.108	0.218	.826
	6. Intrinsische Motivation	.264	.105	2.507	<b>.012</b>

### Multigruppeninvarianz für Länder

Die Untersuchung der Invarianz der faktoriellen Struktur über die Länder (Disziplinen) hinweg war nur für die drei Länder mit einer größeren Anzahl von Befragten (n=445) möglich, nämlich Griechenland (n=177), Portugal (n=156) und Rumänien (n=112). Für diese Analyse wurden für alle Gruppen unabhängige CFAs verwendet; die CFA wurde gleichzeitig auf die drei Gruppen angewendet, um auf Invarianz zu testen. Die Ergebnisse für die drei getesteten Invarianzmodelle sind in Tabelle 9 dargestellt.



**Tabelle 9.** Die Goodness-of-fit Indizes für die Mehrgruppeninvarianz nach Land (Disziplin) für den CTSAS-SF (Faktorenmodell zweiter Ordnung) und SENCTDS (Modell mit sechs korrelierten Faktoren).

	<i>Baseline Modelle</i>	$\chi^2$ (df)	<i>p</i>	<i>RMSEA [90%IC]</i>	<i>CFI</i>	<i>TLI</i>
<b>CTSAS-SF</b>	<i>Gesamtmodell</i>	3639.787 (1695)	<.0001	.051 [.049-.053]	.929	.925
	<i>Portugal</i>	2528.042 (1695)	<.0001	.056 [.052-.061]	.902	.898
	<i>Griechenland</i>	2659.082 (1695)	<.0001	.057 [.056-.061]	.908	.904
	<i>Rumänien</i>	2117.048 (1695)	<.0001	.047 [.040-.053]	.965	.963
	<b><i>Invarianz</i></b>	<b><math>\chi^2</math> (df)</b>	<b><i>p</i></b>	<b><i>RMSEA [90%IC]</i></b>	<b><i>CFI</i></b>	<b><i>TLI</i></b>
	<i>Konfigurative Invarianz</i>	7427.003 (5085)	<.0001	.056 [.053-.058]	.929	.926
	<i>Metrische Invarianz</i>	--	--	--	--	--
	<i>Skalare Invarianz</i>	7922.401 (5769)	<.0001	.050 [.047-.053]	.935	.940
	<b><i>Modellvergleich</i></b>	<b><math>\chi^2</math> (df)</b>	<b><i>p</i></b>	<b><math>\Delta</math>RMSEA</b>	<b><math>\Delta</math>CFI</b>	
	<i>Metrisch vs. Konfigurativ</i>	--	--	--	--	
	<i>Skalar vs. Konfigurativ</i>	849.399 (684)	<.0001	.006	.006	
	<i>Skalar vs. Metrisch</i>	--	--	--	--	
	<b>SENCTDS</b>	<i>Gesamtmodell</i>	528.715 (195)	<.0001	.062 [.056-.068]	.963
<i>Portugal</i>		316.839 (174)	<.0001	.073 [.060-.085]	.959	.950
<i>Griechenland</i>		332.035 (174)	<.0001	.072 [.060-.083]	.955	.946
<i>Rumänien</i>		257.725 (174)	<.0001	.066 [.048-.082]	.973	.967
<b><i>Invarianz</i></b>		<b><math>\chi^2</math> (df)</b>	<b><i>p</i></b>	<b><i>RMSEA [90%IC]</i></b>	<b><i>CFI</i></b>	<b><i>TLI</i></b>
<i>Konfigurative Invarianz</i>		905.093 (522)	<.0001	.070 [.063-.078]	.962	.955
<i>Metrische Invarianz</i>		--	--	--	--	--
<i>Skalare Invarianz</i>		1192.983 (744)	<.0001	.064 [.057-.070]	.956	.963
<b><i>Modellvergleich</i></b>		<b><math>\chi^2</math> (df)</b>	<b><i>p</i></b>	<b><math>\Delta</math>RMSEA</b>	<b><math>\Delta</math>CFI</b>	
<i>Metrisch vs. Konfigurativ</i>		--	--	--	--	
<i>Skalar vs. Konfigurativ</i>		358.373 (222)	<.0001	.006	.006	
<i>Skalar vs. Metrisch</i>		--	--	--	--	

Die Ergebnisse in Bezug auf die skalare Invarianz für die CTSAS-SF-Skala (RMSEA=0,050; CFI=0,935; TLI=0,940) sind besser als die Ergebnisse für das Gesamtmodell (RMSEA=0,051; CFI=0,929; TLI=0,925). Darüber hinaus beträgt der Unterschied zwischen der Konfigurationsinvarianz und der skalaren Invarianz weniger als 0,015 für die RMSEA-Werte und weniger als 0,01 für die CFI-Werte, was die Invarianz zwischen den Ländern/Disziplinen bestätigt. Beim Vergleich der Durchschnittswerte zwischen den drei Ländern wurden für alle sechs latenten Variablen (Fähigkeiten) Unterschiede festgestellt. Die Ausnahme wurde zwischen Griechenland und Portugal für die Fähigkeiten *Interpretation* und *Selbstregulierung* gefunden. Bei den signifikanten Vergleichen sind die Durchschnittswerte in der Gruppe der griechischen und rumänischen Studierenden höher als in der portugiesischen Gruppe. Der faktorielle Vergleich zwischen den Ländern (Portugal bildet die Ausgangsbasis mit dem Mittelwert 0 und der Varianz 1) ist in Tabelle 10 dargestellt.

**Table 10.** Latente Mittelwertunterschiede zwischen Griechenland, Rumänien und Portugal, wobei letzteres als Basiswert mit dem Durchschnitt=0 und der Varianz=1 verwendet wird.

Fähigkeiten		Faktor Mittelwert		SE		Est/SE		p	
		Griechenla nd	Rumänien	Griechenla nd	Rumänien	Griechenla nd	Rumänien	Griechenla nd	Rumänien
CTSAS-SF	1. Interpretation	-.001	.398	.108	.113	-.010	3.514	.992	<.0001
	2. Analyse	.232	.547	.108	.114	2.158	4.792	.031	<.0001
	3. Bewertung	.226	.633	.112	.124	2.020	5.124	.043	<.0001
	4. Schlussfolgerung	.468	.609	.108	.116	4.340	5.255	<.0001	<.0001
	5. Erläuterung	.257	.516	.109	.115	2.362	4.475	.018	<.0001
	6. Selbstregulierung	.076	.335	.105	.116	.720	2.894	.471	.004
<b>Dispositionen</b>									
SENCTDS	1. Reflektion	.475	.371	.124	.129	3.820	2.881	<.0001	.004
	2. Achtsamkeit	-.082	.294	.104	.132	-.791	2.232	.429	.026
	3. Aufgeschlossenheit	-1.054	-.538	.154	.155	-6.834	-3.476	<.0001	.001
	4. Organisation	.044	.184	.117	.148	.376	1.241	.707	.214
	5. Durchhaltevermögen	-.024	.119	.141	.152	-.169	.783	.866	.434
	6. Intrinsische Motivation	-.372	.103	.130	.138	-2.861	.745	.004	.456

Im Gegensatz dazu zeigt die für die SENCTDS-Skala berechnete Multigruppeninvarianz leicht schlechtere Ergebnisse für die Länder im Vergleich zum Gesamtmodell. Die skalaren Invarianzindizes (RMSEA=0,064; CFI=0,956; TLI=0,963) und die geringen Unterschiede beim Vergleich der skalaren vs. konfigurativen Invarianz bestätigen jedoch, dass die faktorielle Struktur in den untersuchten Ländergruppen stabil gehalten wird. Dennoch ist diese Annahme mit Vorsicht zu genießen, denn der Vergleich der faktoriellen Mittelwerte zwischen Griechenland, Rumänien und Portugal (Tabelle 11) zeigt, dass bei der Mehrgruppeninvarianz nur zwei der sechs latenten Faktoren keine Unterschiede aufweisen.

Die geschätzten Korrelationen zeigen, dass die Mittelwerte für die CT-Fähigkeiten in Griechenland und Rumänien höher sind als in Portugal. Was die CT-Dispositionen betrifft, so ergaben die geschätzten Korrelationen höhere Durchschnittswerte in Griechenland und Rumänien als in Portugal, aber die Umkehrung (Portugal weist die höchsten Durchschnittswerte auf) wurde für *Aufgeschlossenheit* festgestellt. Rumänien schnitt bei der *Achtsamkeit* besser ab als Portugal, während Griechenland bei der *Intrinsischen Motivation* schlechter abschnitt als Portugal.

Es war nicht möglich, den Wald-Index zu berechnen, aber Tabelle 11 zeigt die Korrelationen zwischen den Ländern für die CT-Fähigkeiten (CTSAS-SF) und die CT-Dispositionen (SENCTDS). Im Allgemeinen bleibt die faktorielle Korrelation zwischen den Ländern bestehen; die Ausnahmen von diesen Trends sind in Tabelle 11 durch Fettdruck hervorgehoben.

**Tabelle 11.** Faktorkovarianzen nach Ländern für den CTSAS-SF (Fähigkeiten) und den SENCTDS (Dispositionen)

<i>CTSAS-SF</i>	<i>Land</i>	<i>Analyse</i>	<i>Bewertung</i>	<i>Schlussfolgerung</i>	<i>Erläuterung</i>	<i>Selbstregulierung</i>
<i>Interpretation</i>	<i>POR</i>	<b>.910**</b>	<b>.799**</b>	.762**	.715**	<b>.759**</b>
	<i>GRE</i>	<b>.906**</b>	<b>.785**</b>	.790**	.741**	<b>.673**</b>
	<i>ROM</i>	<b>.872**</b>	.801**	.805**	.754**	<b>.741**</b>

<i>CTSAS-SF</i>	<i>Land</i>	<i>Analyse</i>	<i>Bewertung</i>	<i>Schlussfolgerung</i>	<i>Erläuterung</i>	<i>Selbstregulierung</i>
<i>Analyse</i>	<i>POR</i>		<b>.924**</b>	<b>.814**</b>	<b>.786**</b>	<b>.679**</b>
	<i>GRE</i>		<b>.955**</b>	<b>.854**</b>	<b>.810**</b>	<b>.710**</b>
	<i>ROM</i>		<b>.892**</b>	<b>.882**</b>	<b>.848**</b>	<b>.783**</b>
<i>Evaluation</i>	<i>POR</i>			<b>.886**</b>	<b>.763**</b>	<b>.726**</b>
	<i>GRE</i>			<b>.954**</b>	<b>.903**</b>	<b>.735**</b>
	<i>ROM</i>			<b>.926**</b>	<b>.882**</b>	<b>.840**</b>
<i>Bewertung</i>	<i>POR</i>				<b>.842**</b>	<b>.783**</b>
	<i>GRE</i>				<b>.880**</b>	<b>.804**</b>
	<i>ROM</i>				<b>.880**</b>	<b>.885**</b>
<i>Erläuterung</i>	<i>POR</i>					<b>.774**</b>
	<i>GRE</i>					<b>.778**</b>
	<i>ROM</i>					<b>.868**</b>
<i>SENCTDS</i>	<i>Land</i>	<i>Achtsamkeit</i>	<i>Aufgeschlossenheit</i>	<i>Organisation</i>	<i>Durchhaltevermögen</i>	<i>Intrinsische Motivation</i>
<i>Reflektion</i>	<i>POR</i>	<b>.395**</b>	<b>.495**</b>	<b>.469**</b>	<b>.467**</b>	<b>.532**</b>
	<i>GRE</i>	<b>.289**</b>	<b>.443**</b>	<b>.421**</b>	<b>.566**</b>	<b>.545**</b>
	<i>ROM</i>	<b>.416**</b>	<b>.464**</b>	<b>.579**</b>	<b>.608**</b>	<b>.596**</b>
<i>Achtsamkeit</i>	<i>POR</i>		<b>.250**</b>	<b>.072(ns)</b>	<b>.436**</b>	<b>.293**</b>
	<i>GRE</i>		<b>.559**</b>	<b>-.103(ns)</b>	<b>.114(ns)</b>	<b>.074(ns)</b>
	<i>ROM</i>		<b>.388**</b>	<b>-.013(ns)</b>	<b>.516**</b>	<b>.370**</b>
<i>Aufgeschlossenheit</i>	<i>POR</i>			<b>.217**</b>	<b>.500**</b>	<b>.460**</b>
	<i>GRE</i>			<b>-.113(ns)</b>	<b>.203**</b>	<b>.243**</b>
	<i>ROM</i>			<b>.296**</b>	<b>.297**</b>	<b>.287**</b>
<i>Organisation</i>	<i>POR</i>				<b>.395**</b>	<b>.217**</b>
	<i>GRE</i>				<b>.263**</b>	<b>.333**</b>
	<i>ROM</i>				<b>.568**</b>	<b>.532**</b>
<i>Durchhaltevermögen</i>	<i>POR</i>					<b>.758**</b>
	<i>GRE</i>					<b>.656**</b>
	<i>ROM</i>					<b>.834**</b>

*POR- Portugal; GRE – Griechenland; ROM – Rumänien. Alle Korrelationen sind signifikant auf p-Niveau <0.0001*

## 4. Diskussion über den Validierungsprozess des Instruments

Das Instrument, das zur Beobachtung der Veränderungen der CT-Fähigkeiten und CT-Dispositionen verwendet wird, ergibt sich aus der Verbindung zweier unterschiedlicher Selbstauskunftsfragebögen, von denen einer auf die CT-Fähigkeiten abzielt (angepasster CTSAS oder CTSAS-SF) und der andere auf die Beurteilung der CT-Dispositionen (SENCTDS), der in seiner ursprünglichen Form verwendet wurde. Die CTSAS-Kurzform ermöglichte durch die Verkürzung der Zeit zum Ausfüllen des Fragebogens das Ausfüllen beider Fragebögen in weniger als 60 Minuten. Diese Zeitspanne wurde von den Studierenden als angenehm empfunden.

Für die Validierung des endgültigen Fragebogens wurde der erste Satz von Antworten verwendet, die von den in den Pilotkursen eingeschriebenen Studierenden eingereicht wurden (Vortest-Moment). Der Fragebogen wurde zu Beginn der Pilot-CTBACs, die im ersten und zweiten Semester 2021/2022 durchgeführt wurden, bei insgesamt 531 Studierenden der fünf Hochschulen eingesetzt. Die Antworten wurden in der ersten Woche des Kurses, vor Beginn der Lerninterventionen, gesammelt.

Das Alter der Befragten lag zwischen 19 und 58 Jahren (Mittelwert=23,5), wobei 87 % der Studierenden unter 31 Jahre alt waren. In Anbetracht der Tatsache, dass in Europa im Allgemeinen Frauen in den Hochschulen in der Mehrheit sind, insbesondere in Bereichen wie Gesundheitswissenschaften, Kunst, Geisteswissenschaften, Sozialwissenschaften und Bildung, war es nicht überraschend, dass ein höherer Prozentsatz der Befragten weiblich war (75 %; 389 Frauen gegenüber 142 Männern). Der geringe Prozentsatz von Studenten (6,4 %) in den Bereichen Informatik und Technologie, die in der Regel eher von Männern besucht werden, reichte nicht aus, um den prozentualen Unterschied zwischen den Befragten der einzelnen Geschlechter zu verringern.

## CTSAS Validierung

Während der Entwicklung des Think4jobs-Projekts wurde klar, dass das anzuwendende Instrument einen reduzierten Umfang (in Bezug auf die Anzahl der Items) haben sollte, damit es in einer für die Studierenden angenehmen Zeitspanne ausgefüllt werden kann. Der von Nair [24] verfasste CTSAS-Fragebogen hatte insgesamt 115 Items (von 0=nie bis 6=immer), was einer Bearbeitungszeit von 50 Minuten entsprach. Durch die Reduzierung der Anzahl der Items (Eliminierung redundanter Items und Eliminierung von Items mit Ladungsgewichten unter .500) hat der Fragebogen nun eine durchschnittliche Beantwortungszeit von weniger als 30 Minuten, wobei die ursprünglichen sechs Dimensionen (*Interpretation, Analyse, Bewertung, Schlussfolgerung, Erklärung und Selbstregulierung*) mit nur 60 Items beibehalten wurden.

Der Validierungsprozess des CTSAS-SF mit konfirmatorischer Faktorenanalyse führte zu zwei Modellen mit gleichwertigen zufriedenstellenden Anpassungsindizes. Modell 4, das Faktormodell zweiter Ordnung (RMSEA=0,051; TLI=0,924; CFI=0,927) hatte ein Chi-Quadrat/df-Verhältnis von 2,33 und Modell 3, das korrelierte Sechs-Faktoren-Modell, ein Chi-Quadrat/df-Verhältnis von 2,28 (RMSEA=0,049; TLI=0,930; CFI=0,933), was bestätigt, dass beide Modelle eine sehr gute Gesamteignung, interne Konsistenz und Zuverlässigkeit aufweisen. Das Cronbach-Alpha des Gesamtinstruments war ausgezeichnet ( $\alpha = 0,969$ ) [35, 36].

Das Faktormodell zweiter Ordnung (Modell 4) passte besser zu der Annahme, dass kritisches Denken ein komplexes multidimensionales und mehrstufiges Konstrukt ist [37, 38]; diese Annahme wird auch durch die Tatsache gestützt, dass das Modell, das die Hypothese testete, dass alle 60 Items durch einen Faktor erklärt werden (Modell 1), und das bifaktorielle Modell (Modell 5) nicht angemessen zu den Daten passten. Die Faktorenanalyse ergab eine korrelierte Sechs-Faktoren-Struktur: *Interpretation* (9 Items;  $\alpha=0,772$ ), *Analyse* (11 Items;  $\alpha=0,888$ ), *Bewertung* (7 Items;  $\alpha=0,858$ ), *Schlussfolgerung* (13 Items;  $\alpha=0,905$ ), *Erläuterung* (10 Items;  $\alpha=0,853$ ) und *Selbstregulierung* (10 Items;  $\alpha=0,905$ ).

Wenn wir uns auf das Konzept des CT beziehen, müssen wir dessen höherwertige Natur berücksichtigen [39] und bedenken, dass "Unterschiede in der Geschichte eines Lernenden bedeuten können, dass eine Situation, die bei einer Person Denken höherer Ordnung erfordert, bei einer anderen Person nur Denken niedrigerer Ordnung erfordert. Darüber hinaus ist es wahrscheinlich, dass im Unterricht die Vermittlung von grundlegenden und höherwertigen Denkfähigkeiten miteinander verwoben sind". [39].

Ein wichtiger Aspekt, der noch etwas Aufmerksamkeit verdient, ist die Tatsache, dass vier Items (Items 4, 6, 8 und 39) eine faktorielle Ladung unter 0,500 hatten. Diese Situation wurde auch in der Studie von Nair bestätigt, und das Forscherteam beschloss, die vier Items beizubehalten, da wir der Ansicht waren, dass der inhaltliche Gehalt jedes dieser Items für die Charakterisierung wichtiger Aspekte des CT wichtig war.

Die gefundenen Korrelationen zwischen den Fähigkeiten und zwischen den Fähigkeiten und dem integrierten CTSAS-SF-Score waren stark und positiv, was die Existenz einer guten itembezogenen Validität unterstützt und die sehr gute interne Konsistenz und Reliabilität bestätigt. Das Geschlecht hatte keinen Einfluss auf die Datenverteilung, außer bei vier Items (Items 4, 42, 47 und 50). Darüber hinaus behielt der CTSAS-SF seine faktorielle Strukturinvarianz über die Geschlechter hinweg bei, was seine Zuverlässigkeit für beide Geschlechter und Länder bestätigt.

Die CTSAS-SF weist eine sehr gute Validität und Reliabilität auf, und trotz der Verringerung der Anzahl der Items hat die Skala ihre Stabilität beibehalten. Die Skala zeigt ein hohes Potenzial für ihren Einsatz in der Forschung, die die Bewertung von CT bei Hochschulstudenten erfordert, und es zeigte sich, dass sie selbst bei der Anwendung in einem multinationalen Kontext bei Studenten in fünf sehr unterschiedlichen Hochschulprogrammen gute Ergebnisse erzielte.

## SENCTDS Validierung

Der SENCTDS-Validierungsprozess mit konfirmatorischer Faktorenanalyse ergab nur ein Modell mit zufriedenstellenden Anpassungsindizes - das Modell 3 mit sechs korrelierten Faktoren (RMSEA=0,054; TLI=0,974; CFI=0,969) mit einem Chi-Quadrat/df-Verhältnis von 2,57. Diese Werte belegen eine sehr gute Gesamtauglichkeit, interne Konsistenz und Zuverlässigkeit des Modells. Das SENCTDS-Instrument wies außerdem ein hohes Cronbach-Alpha ( $\alpha = 0,842$ ) auf, was auf eine starke interne Konsistenz des Instruments hindeutet [40]. Darüber hinaus wiesen die einzelnen Dimensionen der mit SENCTDS erfassten CT-Dispositionen akzeptable bis gute Cronbachs Alpha-Werte auf [35, 36, 40]: *Reflektion* (3 Items;  $\alpha=.796$ ), *Achtsamkeit* (4 Items;  $\alpha=0,853$ ), *Aufgeschlossenheit* (4 Items;  $\alpha=0,773$ ), *Organisation* (3 Items;  $\alpha=0,772$ ), *Durchhaltevermögen* (3 Items;  $\alpha=0,792$ ) und *Intrinsische Motivation* (4 Items;  $\alpha=0,842$ ). Diese Koeffizienten stützen die Schlussfolgerung, dass die Konstrukte die beabsichtigten Dimensionen messen, was darauf hindeutet, dass jede der sechs Skalen unabhängig verwendet werden kann, wenn eine bestimmte Dimension gesondert bewertet werden muss [41].

Auch wenn die faktoriellen Kovarianzen in SENCTDS ein eher einheitliches Verhalten zwischen Frauen und Männern vermuten lassen, hat die Invarianzanalyse gezeigt, dass Männer im Vergleich zu Frauen höhere Werte in den Bereichen *Achtsamkeit* und *Intrinsische Motivation*, aber niedrigere Werte in der Kategorie *Organisation* erzielen. Die Invarianzanalyse für Land/Disziplin bestätigte, dass die faktorielle Struktur zwischen den Ländern beibehalten wurde, auch wenn nur zwei der sechs Dispositionsfaktoren keine Unterschiede in den faktoriellen Mittelwerten zwischen den Ländern aufweisen. Auch wenn dieses Verhalten auf bestehende Unterschiede zwischen den Ländern in Bezug auf das Skalenverhalten hinweist, können wir die Auswirkungen der unausgewogenen Länderrepräsentativität unter den Befragten nicht ausschließen [42].

## TEIL II – DISZIPLINÜBERGREIFENDE ANALYSE DER CTBACS IMPLEMENTATION

Die Bewertung der Wirksamkeit des Unterrichts ist von entscheidender Bedeutung, insbesondere wenn neue Unterrichtsstrategien entwickelt werden, die auf die Entwicklung der Studierenden in einem bestimmten Bereich abzielen (sei es die kognitive Entwicklung oder die Entwicklung der Einstellung) oder in einem Kurs oder fächerübergreifend eingeführt werden. Die Evaluierung der Ergebnisse solcher Interventionen fördert und unterstützt Lehrplanänderungen und sollte nicht übersehen werden.

In Ermangelung eines Standardinstruments zur Bewertung der Lerneffektivität greifen Lehrkräfte und Einrichtungen häufig auf eine Vortest/Nachtest-Bewertung der angestrebten Lernziele in einem Kurs (kurzfristiges Lernen und Längsschnittbewertung) oder im Rahmen eines Bildungsprogramms und der Qualität der vermittelten Ausbildung (langfristiges Lernen und transversale Bewertung) zurück [43]. Der Vortest/Nachtest-Ansatz ermöglicht es, die Basiskompetenz der Studierenden zu Beginn des Unterrichts zu ermitteln und die Verbesserungen am Ende des Unterrichts zu messen [43]. Obwohl dieser Vortest/Nachtest-Ansatz häufig verwendet wird, zielt er in vielen Fällen auf die Bewertung des kognitiven Wissens und weniger häufig auf die Soft Skills oder die mit den Studierenden trainierten Dispositionen ab.

Außerdem würde die regelmäßige Anwendung des Vortest/Nachtest-Ansatzes in allen Studiengängen als Instrument für eine solide Überwachungsstrategie es ermöglichen, das Wissen und die Kompetenzen der Studierenden zu verfolgen und methodisch anzupassen, um entstehende Lücken, die sich aus der ständigen technologischen Entwicklung oder den kritischen Herausforderungen in den meisten Berufen ergeben, abzumildern.

Die verfügbare Literatur über die Bewertung der Umsetzung neuer pädagogischer Maßnahmen deutet jedoch auf einige Schwierigkeiten bei der Bewertung der Wirksamkeit hin, nämlich die Verwendung mehrerer pädagogischer Maßnahmen (in Bezug auf Form und

Dauer), eine unzureichende Kategorisierung der pädagogischen Maßnahmen, das Fehlen eines Pretests zur Ermittlung der Ausgangssituation, die Verwendung nicht validierter Instrumente, die Erfassung der von den Studierenden wahrgenommenen Fortschritte oder die geringe Anzahl von Studierenden in der Studie. Außerdem sind die meisten Studien quasi-experimentell, da es schwierig ist, eine Kontrollgruppe innerhalb desselben Kurses zu bilden [17].

In diesem Abschnitt zielt der Bericht darauf ab, mögliche kurzfristige Veränderungen nach der Durchführung der zwölf CT-gemischten Lehrpläne, die in IO3 [1] vorgestellt wurden, zu bewerten. In dieser Evaluierung wurde der Schwerpunkt auf die Verbesserung der CT-Fähigkeiten und CT-Dispositionen innerhalb eines Kurses gelegt und nicht auf den Erwerb von kognitivem Wissen, das für diesen Kurs spezifisch ist, auch wenn die CT-Dimensionen im spezifischen Kontext eines Kurses oder einer Disziplin verankert wurden. Um diese Veränderungen zu analysieren, wurden die (für diese Population validierten) Skalen CTSAS-SF und SENCTDS vor und nach den pädagogischen Interventionen verwendet.

Die im Laufe dieses Prozesses gesammelten Daten werden anhand der Veränderungen analysiert, die sich aus dem Vergleich der Studierendenergebnisse zu Beginn und am Ende der Pilotkurse ergeben (länderübergreifende Vergleiche), aber auch innerhalb der einzelnen Länder, da es aufgrund der vorhandenen Störvariablen schwierig ist, sinnvolle Vergleiche zwischen den Ländern anzustellen, nämlich:

- Die Unterschiede in den in der Studie vertretenen Disziplinen (eine Disziplin - ein Land);
- Die Unterschiede bei den Studierenden, die sich aus dem individuellen (kulturellen oder religiösen, u.a.) und disziplinären Hintergrund ergeben, sowie die Unterschiede in ihrem Engagement für die Aktivitäten;
- Die Unterschiede bei den durchgeführten Maßnahmen;

- Der Einsatz unterschiedlicher Lehrkräfte in verschiedenen Kursen, was sich auf die zwischenmenschlichen Beziehungen und die Art und Weise, wie das Lernen erreicht wird, auswirken kann.

## 1. Methodologie

Das Projekt schlug vor, die mit dem CT verbundenen Fähigkeiten und Neigungen zu messen, um den Erfolg der eingeführten CTBACs in Pilotkursen in Wirtschaftsinformatik (Deutschland), Lehrerbildung (Griechenland), Englisch als Fremdsprache (Litauen), Veterinärmedizin (Portugal) und Wirtschaftswissenschaften (Rumänien) zu messen. Der eingereichte Vorschlag sah vor, die Instrumente zur Bewertung der CT-Fähigkeiten und CT-Dispositionen dreimal während jeder Kursdauer einzusetzen sowie einen experimentellen Ansatz zu verfolgen, indem eine Versuchs- und eine Kontrollgruppe gebildet wurden.

Die Instrumente CTSAS-SF und SENCTDS wurden verwendet, um die Veränderungen der CT-Fähigkeiten bzw. -Dispositionen zu ermitteln. Wir sahen uns jedoch mit einigen Einschränkungen konfrontiert, um den vorgeschlagenen Plan zu erfüllen. Zum einen wurden die Pilotkurse in den meisten Ländern/Disziplinen (Deutschland, Litauen, Portugal und Rumänien) in dem für die Einführung der CTBACs vorgesehenen Zeitraum nur einmal angeboten. Zum anderen machen die Studierenden geltend, dass die Bildung einer Kontrollgruppe innerhalb derselben Lehrplaneinheit zu einer ungleichen Lernqualität führen könnte, da sie die Unterschiede in den Lerninterventionen als Diskriminierung ansehen, wobei die Versuchsgruppe einen vermeintlichen Vorteil erhält. Folglich wurden alle im Pilotkurs eingeschriebenen Studierenden den vorgeschlagenen Interventionen unterzogen. Für diese Länder entwickelte sich die Studie zu einer Quasi-Experimentalsituation.

Eine weitere Einschränkung war der Verlust von Studierenden während der drei Zeitpunkte der Anwendung der CT-Bewertungsinstrumente. Außerdem haben einige von ihnen einen der Fragebögen nicht ausgefüllt. Infolgedessen legten am Ende der Studie nur sehr wenige

Studierende die drei geforderten ausgefüllten Fragebögen vor; oft fehlte der mittlere. Daher beschloss die Partnerschaft, nur die Daten zu verwenden, die beim Vortest (zum Zeitpunkt Null des Kurses) und beim Nachtest (der am Ende des Faches ausgefüllt wurde) gesammelt wurden, um den Erfolg der pädagogischen Interventionen zu bewerten.

Die Links zu den zusammengeführten Fragebögen CTSAS-SF und SENCTDS wurden, wie bereits erläutert, über die Moodle-Seite jedes Kurses an die Studierenden weitergegeben. Neben der englischen Version wurden vier Übersetzungen der Fragebögen auf Deutsch, Griechisch, Portugiesisch und Rumänisch zur Verfügung gestellt (siehe Abschnitt I für Details).

Alle Studierenden unterzeichneten eine Einverständniserklärung, in der sie erklärten, dass sie die Ziele des Projekts verstanden haben, dass sie jederzeit aus der Studie aussteigen können, ohne eine Strafe zahlen zu müssen, und dass alle Daten gemäß den europäischen Vorschriften vertraulich behandelt und anonymisiert werden.

## Teilnehmer

Für die aktuelle Studie wurde eine nicht-randomisierte Zufallsstichprobe von Studierenden verwendet, die in den Pilotkursen eingeschrieben waren, in denen die CTBACs eingeführt wurden. Von den 609 Studierenden, die an den Pilotaktivitäten teilnahmen, beantworteten nur 87,2 % die Fragebögen. Von den 531 Studierenden, die zu Beginn des Kurses den Vortest beantworteten, antworteten nur 63,1 % der Studierenden auf den Nachtest-Fragebogen, der am Ende des Kurses ausgefüllt wurde. Die gepaarten Fragebögen erreichen eine Repräsentativität von 55 % der an dem Projekt beteiligten Population. Die Grundgesamtheit, die zur Bewertung des Erfolgs der Pilotmaßnahmen herangezogen wurde, setzt sich folglich aus 335 Studierenden zusammen (Tabelle 12); 258 (77 %) waren weiblich und 77 (23 %) waren männlich. Die unausgewogene Verteilung der Geschlechter folgte dem in Teil I beschriebenen Muster. Das Durchschnittsalter betrug 23,62 Jahre ( $sd = 7,50$ ; Spanne 18 -58).

Die Versuchsgruppe umfasste 286 Studierende (85,4 %), während 49 Studierende (14,6 %) die Kontrollgruppe bildeten. Die Studierenden vertraten fünf Länder; die meisten von ihnen stammten aus Portugal (32,5%), Griechenland (30,7%) und Rumänien (24,2%). Studierende aus Deutschland (6,6%) und Litauen (6,0%) stellten nur 12,6% der Teilnehmer (Tabelle 12). Die Kontrollgruppe setzte sich aus griechischen (n= 49) und portugiesischen (n=9) Studierenden zusammen.

**Tabelle 12.** Verteilung der Studierendenpopulation nach Disziplin und Kurs (CTBACs und Kontrolle) und ihre Repräsentativität für die Datenbank der gepaarten Fragebögen.

Land	Programm/Disziplin	Modul	Anzahl der Studierenden		% der gesamten gepaarten Fragebögen
			Vortest	Nachtest	
Germany	Wirtschaftsinformatik	Design Patterns	11	8	2,4
		Innovation Management	10	4	1.2
		Economic Aspects of Industrial Digitalization	10	8	2.4
		Wissenschaftliches Seminar	10	2	.6
Griechenland	Pädagogik	Didaktik der Biologie	46	22	6.6
		Didaktik der Naturwissenschaft	111	74	22.1
		Didaktik der Umweltkunde	20	7	2.1
Litauen	Internationale Beziehungen und Politikwissenschaft	Englisch als Fremdsprache	52	20	6.0
Portugal	Veterinärmedizin	Bildgebung	55	36	10.7
		Deontologie	52	42	12.5
		Gynäkologie, Andrologie und Geburtshilfe	41	22	6.6
		Curriculare Praktika	12	9	2.7
Rumänien	Betriebs- und Volkswirtschaft	Wirtschaftskommunikation	35	31	9.3
		Pädagogik und Didaktik der	40	32	9.6
		Finanzbuchhaltung	26	18	5.4
		Virtuelle Lernumgebungen in der Wirtschaftswissenschaft			
<b>Gesamt</b>			<b>531</b>	<b>335</b>	<b>100</b>

### Charakterisierung der Experimentalgruppe

Die Experimentalgruppe umfasste 286 Studenten, von denen 76,2% (n = 218) weiblich und 23,8% (n = 68) männlich waren. Das Durchschnittsalter betrug 23,88 Jahre (sd = 7,62; Spanne 18 - 54). 35% (n = 100) der Studierenden kamen aus Portugal, gefolgt von 28,3% (n = 81) aus Rumänien und 22,0% (n = 63) aus Griechenland. Studierende aus Deutschland und Litauen

stellen 7,7% (n = 22) bzw. 7,0% (n = 20) der Studierenden. Die Verteilung der Studierenden nach Fachrichtung und Studiengang ist in Tabelle 13 dargestellt, die Altersverteilung in Tabelle 14.

**Tabelle 13.** Verteilung der Versuchsgruppe der Studierenden nach Fachrichtung und Modul

Programm/Disziplin	Module	n	%
Wirtschaftsinformatik (n=22)	Design Patterns	8	2.8
	Innovation Management	4	1.4
	Economic Aspects of Industrial Digitalization	8	2.8
	Wissenschaftliches Seminar	2	.7
Pädagogik (n=63)	Didaktik der Biologie	22	7.7
	Didaktik der Naturwissenschaft	34	11.9
	Didaktik der Umweltkunde	7	2.4
Internationale Beziehungen und Politikwissenschaft (n=20)	Englisch als Fremdsprache	20	7.0
Veterinärmedizin (n=100)	Bildgebung	36	12.6
	Deontologie	42	14.7
	Gynäkologie. Andrologie und Geburtshilfe	22	7.7
Betriebs- und Volkswirtschaft (82)	Wirtschaftskommunikation	31	10.8
	Pädagogik und Didaktik der Finanzbuchhaltung	32	11.2
	Virtuelle Lernumgebungen in der Wirtschaftswissenschaft	18	6.3
<b>Gesamt</b>		<b>286</b>	<b>100.0</b>

**Tabelle 14.** Altersverteilung in der Experimental- und Kontrollgruppe

	<i>Experimentalgruppe</i>		<i>Kontrollgruppe</i>		
	<i>N</i>	<i>Prozentualer Anteil</i>	<i>N</i>	<i>Prozentualer Anteil</i>	
<b>Alter (Jahre)</b>					
	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
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
<b>Gesamt</b>	<b>286</b>	<b>100</b>	<b>49</b>	<b>100,0</b>

### **Charakterisierung der Kontrollgruppe**

Die Kontrollgruppe bestand aus 49 Befragten, davon 40 aus Griechenland (Fachrichtung Pädagogik – Modul Naturwissenschaft) und neun aus Portugal (Fachrichtung Veterinärmedizin - Curriculare Praktika). Das Durchschnittsalter betrug 22,10 Jahre (sd =6,61; Spanne 19 -58). Vierzig Studierenden waren weiblich und neun waren männlich.

### **Datenanalyse**

In der CTSAS-SF-Skala setzen sich die Fähigkeiten (Dimensionen) aus zwei bis drei Teilfähigkeiten (Subdimensionen) zusammen, die wiederum aus einer unterschiedlichen Anzahl von Items zusammengesetzt sind. Zur Berechnung der Punktzahl für jede Teilkompetenz wurde der Durchschnitt der Punktzahlen für die Items der entsprechenden Teildimension verwendet. Zur Berechnung der Punktzahl für eine bestimmte Fähigkeit wurde die Summe der für jede entsprechende Subdimension erhaltenen Werte verwendet. Die Wertebereiche für die Fähigkeiten und Teilfähigkeiten sind in Tabelle 15 aufgeführt.

In Anlehnung an die SENCTDS-Skala wurden vor der Berechnung des Mittelwerts für jede Dimension die Punkte 4 bis 11 umgedreht. Die Punktzahl für jede Dimension (Disposition) der Skala wurde als Mittelwert errechnet. Die Wertebereiche für die Dimensionen des SENCTDS-

Instruments sind in Tabelle 15 aufgeführt. Die integrierten Punktzahlen für Fähigkeiten und Dispositionen stellen die Summe aller Dimensionen der Skala dar. Nach der Berechnung der Ergebnisse für jede Dimension und Subdimension wurde die statistische Analyse durchgeführt.

**Tabelle 15.** Wertebereich für jede Dimension und Subdimension der Skalen SCTSAS-SF und SENCTDS.

Fähigkeiten	Wertebereich	Teilfähigkeiten	Items	Wertebereich (Mittelwert)
<b>Interpretation</b>	0 – 18	<i>Kategorisierung</i>	1 - 3	0 – 6
		<i>Bedeutung klären</i>	6 - 9	0 – 6
		<i>Bedeutung entschlüsseln</i>	4, 5	0 – 6
<b>Analyse</b>	0 – 18	<i>Erkennen von Argumenten</i>	15, 16	0 – 6
		<i>Analysieren von Argumenten</i>	17 - 20	0 – 6
		<i>Prüfung von Ideen</i>	10 - 14	0 – 6
<b>Bewertung</b>	0 – 12	<i>Behauptung bewerten</i>	21, 22	0 – 6
		<i>Bewertung von Argumenten</i>	23 - 27	0 – 6
<b>Schlussfolgerung</b>	0 – 18	<i>Ziehen von Schlussfolgerungen</i>	36 - 40	0 – 6
		<i>Vermutungen über Alternativen anstellen</i>	31 - 35	0 – 6
		<i>Beweise abfragen</i>	28 - 30	0 – 6
<b>Erläuterung</b>	0 – 18	<i>Ergebnisse darlegen</i>	41, 42	0 – 6
		<i>Verfahren rechtfertigen</i>	43, 44	0 – 6
		<i>Argumente vorbringen</i>	45 - 50	0 – 6
<b>Selbstregulierung</b>	0 – 12	<i>Selbstüberprüfung</i>	51 - 57	0 – 6
		<i>Selbstkorrektur</i>	58 - 60	0 – 6
<b>Integrierter Wert</b>	0 – 96		1 - 60	
Dispositionen		Items	Wertebereich (Mittelwert)	
<b>Reflektion</b>		1 - 3	1 - 7	
<b>Achtsamkeit</b>		4 - 7 (reverted)	1 - 7	
<b>Aufgeschlossenheit</b>		8 - 11 (reverted)	1 - 7	
<b>Organisation</b>		12 - 14	1 - 7	

<i>Durchhaltevermögen</i>	15 - 17	1 - 7
<i>Intrinsische Motivation</i>	18 - 21	1 - 7
<i>Integrierter Wert</i>	1 - 21	6 - 42

## Statistische Analyse

Die demografischen Daten (Geschlecht, Alter und Land) wurden mit Hilfe der deskriptiven Statistik analysiert. Der T-Test für unabhängige Stichproben, die einseitige ANOVA und die Pearson-Korrelation wurden verwendet, um die Ausgangsunterschiede für Geschlecht, Modul und Alter zu testen.

Der T-Test für gepaarte Stichproben wurde verwendet, um die allgemeinen Veränderungen in den CT-Fähigkeiten und C T-Dispositionen zu bewerten, die zu zwei verschiedenen Zeitpunkten (vor und nach der letzten pädagogischen Intervention) gemessen wurden. Eine einseitige ANOVA wurde angewandt, um Länderunterschiede bei den Veränderungen vor und nach der pädagogischen Intervention zu testen.

Um die Wirkung der verschiedenen Strategien nach Ländern zu analysieren, wurde eine GLM-Univariate ANOVA mit dem Ergebnis nach der Intervention als abhängige Variable, den verschiedenen Strategien als unabhängige Variable und dem Ergebnis vor der Intervention als Kovariate verwendet.

Alle Analysen wurden mit IBM SPSS Statistics 26 durchgeführt. Alle Tests wurden mit einem statistischen Signifikanzniveau von  $p \leq 0,05$  und einem Konfidenzintervall von 95% durchgeführt.

## 2. Ergebnisse

### Prüfung auf Unterschiede bei den Ausgangswerten in der Versuchsgruppe

Das Vorhandensein von Ausgangsunterschieden zwischen den Geschlechtern und Ländern für die Versuchsgruppe wurde sowohl für die CT-Fähigkeiten als auch für die Dispositionen getestet. Es wurden keine signifikanten Unterschiede zwischen den Geschlechtern bei den Pretest-Ergebnissen für die integrierte Bewertung der CT-Fähigkeiten sowie für die einzelnen Fähigkeiten und Teilfähigkeiten festgestellt. Dennoch wurde ein signifikanter Restunterschied in der Teilfertigkeit "*Erläuterung: Präsentation von Argumenten*" ( $p=0,055$ ) festgestellt, wo die männlichen Studierenden einen höheren Mittelwert im Vergleich zu den weiblichen Studierenden (3,85 vs. 3,59) aufwiesen.

Signifikante Unterschiede zwischen männlichen und weiblichen Studierenden wurden in den Dimensionen *Reflektion* ( $p = 0,022$ ), *Aufgeschlossenheit* ( $p = 0,018$ ) und *Organisation* ( $p \leq 0,0001$ ) sowie in der integrierten Bewertung der Dispositionen ( $p = 0,013$ ). In allen vier Situationen wiesen Frauen im Vergleich zu Männern höhere Durchschnittswerte auf (*Reflektion*: 5,81 vs. 5,52; *Aufgeschlossenheit*: 5,52 vs. 5,13; *Organisation*: 5,13 vs. 4,25; Integrierte Bewertung der Dispositionen: 31,24 vs. 29,75).

Signifikante Unterschiede zwischen den Ländern wurden bei allen CT-Fähigkeiten, Teilfähigkeiten und der integrierten Bewertung festgestellt, mit Ausnahme der Teilfähigkeiten: *Interpretation - Bedeutung erklären* ( $p = 0,396$ ), *Erklärung - Verfahren begründen* ( $p = 0,052$ ) und *Erklärung - Argumente vorbringen* ( $p = 0,131$ ). Im Allgemeinen erzielten die rumänischen Studierenden bei allen Fertigkeiten und Teilfertigkeiten sowie bei der integrierten Bewertung die höheren Durchschnittswerte. Die litauischen Studierenden erzielten die niedrigsten Durchschnittswerte. Aufgrund der kleinen Stichprobe litauischer und deutscher Studierender ist bei der Interpretation dieser Ergebnisse jedoch Vorsicht geboten.

Hinsichtlich der Skala der CT-Dispositionen wurden keine signifikanten Unterschiede in den Durchschnittswerten zwischen den fünf Ländern festgestellt, mit Ausnahme der Dimension Aufgeschlossenheit, deren Durchschnittswerte in Tabelle 16 dargestellt sind.

**Tabelle 16.** Durchschnittliche Punktzahl pro Land in der Dimension Aufgeschlossenheit

Land	n	Mean	SD
Portugal	100	5.91	1.08
Rumänien	81	5.18	1.28
Griechenland	63	5.06	1.19
Litauen	20	5.51	.80
Deutschland	22	5.13	1.18
<b>Gesamt</b>	<b>286</b>	<b>5.43</b>	<b>1.21</b>

### **Beziehung zwischen Alter und CT-Fähigkeiten und -Dispositionen**

Es wurde ein statistisch positiver Zusammenhang zwischen dem Alter der Studierenden und der integrierten Bewertung der CT-Fähigkeiten ( $r = 0,193$ ;  $p = 0,001$ ) und jeder CT-Fähigkeit mit Ausnahme der Bewertung ( $r = 0,097$ ;  $p = 0,102$ ) festgestellt. Ein positiver Zusammenhang wurde auch zwischen dem Alter und den CT-Teilfähigkeiten festgestellt, mit Ausnahme der in Tabelle 17 aufgeführten Teilfähigkeiten.

Ein positiver signifikanter Zusammenhang wurde für 3 der 6 Dimensionen der Dispositionsskala und für den integrierten Score gefunden; die Ausnahmen waren *Achtsamkeit* ( $r = 0,035$ ;  $p = 0,558$ ), *Aufgeschlossenheit* ( $r = -0,04$ ;  $p = 0,948$ ) und *Organisation* ( $r = 0,106$ ;  $p = 0,073$ ) (Tabelle 17).

**Tabelle 17.** CT-Fähigkeiten und -Teilfähigkeiten sowie Dispositionen in der Versuchsgruppe nicht vom Alter der Studierenden beeinflusst

Fähigkeiten and Teilfähigkeiten	r	p
<i>Interpretation_Entschlüsselung der Bedeutung</i>	.083	.162
<i>Bewertung_Anspruch beurteilen</i>	.061	.305

<i>Erläuterung_Begründungsverfahren</i>	.097	.101
<i>Erläuterung_Argumente vorbringen</i>	.076	.198
<i>Selbstregulierung_Selbstkorrektur</i>	.096	.107
<b>Dispositionen</b>		
<i>Achtsamkeit</i>	.035	.558
<i>Aufgeschlossenheit</i>	-.004	.948
<i>Organisation</i>	.106	.073

### Einfacher Gesamtvergleich vor und nach dem Test für die Experimentalgruppe

Der Gesamtvergleich der Ergebnisse vor und nach dem Test (ohne Berücksichtigung der bereits erwähnten anfänglichen Unterschiede) bestätigt einen Zuwachs an CT-Fähigkeiten in der Versuchsgruppe der Studierenden (Tabelle 16). Allerdings ist der Zuwachs bei den Dispositionen nicht so offensichtlich; nur die Dimension *Organisation* zeigte eine signifikante Verbesserung. Wie aus Tabelle 18 hervorgeht, nahm die Disposition *Aufgeschlossenheit* zwischen den Ergebnissen vor und nach dem Test sogar signifikant ab.

**Tabelle 18.** Vergleich der CT-Fähigkeiten und CT-Dispositionen in der Gruppe der Studierenden, die an den CTBACs teilgenommen haben, bei den Messungen vor und nach dem Test

Fähigkeiten	Mittelwert	SD	Differenz von Mittelwerten	t	p
<i>POST_ Interpretation</i>	13.67	2.20	1.01	7.357	≤.0001
<i>PRE_ Interpretation</i>	12.66	2.31			
<i>POST_ Bewertung</i>	8.85	1.72	.73	7.134	≤.0001
<i>PRE_ Bewertung</i>	8.13	1.85			
<i>POST_ Analyse</i>	13.08	2.52	1.12	7.650	≤.0001
<i>PRE_ Analyse</i>	11.96	2.74			
<i>POST_ Schlussfolgerung</i>	13.42	2.51	.95	6.779	≤.0001
<i>PRE_ Schlussfolgerung</i>	12.47	2.63			
<i>POST_ Erläuterung</i>	12.88	2.50	1.42	9.688	≤.0001
<i>PRE_ Erläuterung</i>	11.46	2.66			

Fähigkeiten	Mittelwert	SD	Differenz von Mittelwerten	t	p
<i>POST_ Selbstregulierung</i>	9.17	1.77	.61	6.151	<b>≤.0001</b>
<i>PRE_ Selbstregulierung</i>	8.57	1.93			
<i>POST_ INTEGRIERTER WERT</i>	71.09	11.79	5.85	9.705	<b>≤.0001</b>
<i>PRE_ INTEGRIERTER WERT</i>	65.24	12.02			
DISPOSITIONEN	Mittelwert	SD	Differenz von Mittelwerten e	t	p
<i>POST_ Reflektion</i>	5.84	.88	.10	1.766	<b>.079</b>
<i>PRE_ Reflektion</i>	5.74	.89			
<i>POST_ Achtsamkeit</i>	3.94	1.46	-.06	-.805	.422
<i>PRE_ Achtsamkeit</i>	4.00	1.42			
<i>POST_ Aufgeschlossenheit</i>	5.23	1.43	-.20	-2.636	<b>.009</b>
<i>PRE_ Aufgeschlossenheit</i>	5.43	1.21			
<i>POST_ Organisation</i>	5.10	1.32	.18	2.568	<b>.011</b>
<i>PRE_ Organisation</i>	4.92	1.41			
<i>POST_ Durchhaltevermögen</i>	5.47	1.17	.07	1.182	.238
<i>PRE_ Durchhaltevermögen</i>	5.40	1.12			
<i>POST_ Intrinsische Motivation</i>	5.49	1.08	.10	1.712	<b>.088</b>
<i>PRE_ Intrinsische Motivation</i>	5.39	1.07			
<i>POST_ INTEGRIERTER WERT</i>	31.07	4.65	.18	.854	.394
<i>PRE_ INTEGRIERTER WERT</i>	30.88	4.33			

## Fächerübergreifende Vergleiche für die Experimentalgruppe

In diesem Projekt repräsentiert jedes Land eine bestimmte Disziplin. Für die Zwecke dieser Analyse überschneiden sich daher die Begriffe "Land" und "Disziplin" und werden austauschbar verwendet.

Die Länder unterschieden sich signifikant in der mittleren Differenz zwischen den Ergebnissen vor und nach dem Test für die integrierte Bewertung der CT-Fähigkeiten ( $F = 2,993$ ;  $df = 4$ ;  $p = 0,019$ ), aber nicht für die integrierte Bewertung der CT-Dispositionen ( $F = 1,808$ ;  $df = 4$ ;  $p = 0,127$ ). Hinsichtlich der CT-Fähigkeiten zeigten die Ergebnisse der Nachtests, dass diese

Unterschiede vor allem bei den rumänischen und litauischen Studierenden zu finden waren ( $p < 0,050$ ); der Zuwachs war bei den ersteren geringer und bei den letzteren am höchsten.

Als die Analyse der Zuwächse nur mit den Daten der drei am stärksten vertretenen Länder (Portugal, Griechenland und Rumänien) wiederholt wurde, weil die geringe Zahl der Befragten aus Litauen und Deutschland die Aussagekraft der Analyse beeinträchtigte, unterschieden sich die Ergebnisse der Nachtests signifikant von denen der Vortests sowohl bei den integrierten Fähigkeiten ( $F = 3,312$ ;  $df = 2$ ;  $p = 0,038$ ) als auch bei den Dispositionen ( $F = 3,224$ ;  $df = 2$ ;  $p = 0,042$ ) (Tabelle 19). Der Zuwachs an CT-Fähigkeiten und -Dispositionen war bei den portugiesischen Studierenden am höchsten, während die rumänischen Studierenden den geringsten Zuwachs aufwiesen (Tabelle 19). Die Post-hoc-Tests zeigten, dass es signifikante Unterschiede zwischen den portugiesischen und den rumänischen Studierenden gab, nicht aber zwischen den portugiesischen und den griechischen Studierenden oder zwischen den rumänischen und den griechischen Studierenden.



**Tabelle 19.** Zuwächse in den integrierten Ergebnissen für CT-Fähigkeiten und CT-Dispositionen in der experimentellen Gruppe von Studierenden, die an CTBACs teilnehmen, in den drei Ländern mit einer größeren Teilnehmerzahl.

		<b>N</b>	<b>Mittelwert*</b>	<b>Std. Deviation</b>	<b>Minimum</b>	<b>Maximum</b>	<b>F (df = 2)</b>	<b>p</b>
<b>Anderungen bei Fähigkeiten</b>	<b>Portugal</b>	100	7.19	9.97	-1.14	1.75	3.312	<b>.038</b>
	<b>Rumänien</b>	81	3.43	10.49	-1.85	1.79		
	<b>Griechenland</b>	63	6.13	8.99	-.91	1.88		
	<b>Gesamt</b>	244	5.67	10.20	-1.85	1.88		
<b>Anderungen bei Dispositionen</b>	<b>Portugal</b>	100	.92	3.50	-1.61	1.72	3.224	<b>.042</b>
	<b>Rumänien</b>	81	-.25	3.88	-1.82	1.46		
	<b>Griechenland</b>	63	-.30	3.36	-1.42	1.11		
	<b>Gesamt</b>	244	.21	3.63	-1.82	1.72		

\* Der Mittelwert stellt die durchschnittliche Differenz zwischen den Ergebnissen nach dem Test und den Ergebnissen vor dem Test in jedem Land dar. Je höher und positiver der Mittelwert ist, desto größer ist die Veränderung der Punktzahl nach der Intervention.

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



## Veränderungen bei den CT-Fähigkeiten und CT-Dispositionen pro Disziplin und Land

### Deutschland

Die deutsche Stichprobe bestand aus 22 Studenten, die im Fachbereich Wirtschaftsinformatik eingeschrieben waren, mit einem Durchschnittsalter von 24,05 Jahren ( $sd = 5,28$ ; Bereich = 19 - 38). 81,8% der Teilnehmer waren männlich und 18,2% waren weiblich. Die Studenten nahmen an dieser Studie im Rahmen von vier Modulen teil: Design Patterns ( $n=8$ ); Economic Aspects of Industrial Digitalization ( $n=8$ ); Innovationsmanagement ( $n=4$ ) und Wissenschaftliches Seminar ( $n=2$ ). Tabelle 20 zeigt die deskriptiven Ergebnisse für die einzelnen Skalen und Subskalen.

**Tabelle 20.** Deskriptive Statistiken für CT-Fähigkeiten und CT-Dispositionen der deutschen Stichprobe zu Beginn der Pilotmodule

	Vorstest Werte	Mittelwert	Std. Abweichung	Minimum	Maximum
<b>Fähigkeiten</b>	Interpretation	11.71	2.68	7.08	16.58
	Analyse	8.13	1.83	3.30	11.10
	Bewertung	10.80	3.26	4.20	16.20
	Schlussfolgerung	12.23	2.66	6.67	17.47
	Erläuterung	11.34	2.70	6.17	16.50
	Selbstregulierung	8.32	1.88	3.86	11.86
	<b>Integrierter Wert für CT-Fähigkeiten</b>	<b>62.54</b>	<b>13.41</b>	<b>32.86</b>	<b>85.14</b>
<b>Dispositionen</b>	Reflektion	5.64	.66	4.67	6.67
	Achtsamkeit	4.10	1.38	2.00	6.25
	Aufgeschlossenheit	5.13	1.18	2.50	7.00
	Organisation	4.82	1.28	1.67	6.67
	Durchhaltevermögen	5.45	.96	3.67	7.00
	Intrinsische Motivation	5.65	.89	3.50	7.00
	<b>Integrierter Wert für CT-Dispositionen</b>	<b>30.78</b>	<b>3.88</b>	<b>25.25</b>	<b>38.92</b>

Die Überprüfung der vorherigen Gleichheit zwischen den Gruppen nach Alter oder Kurs wurde nicht durchgeführt, da die Gruppen nicht ausgewogen waren und die Repräsentativität der Gruppen für die Variable Kurs gering war. Das Alter wirkte sich weder bei den CT-Fähigkeiten noch bei den CT-Dispositionen auf die Vorabgleichung aus. Die Vergleiche zwischen Vor- und Nachtests ergaben insgesamt einen positiven Effekt der Intervention in zwei Fertigkeiten: "Interpretation" und "Analyse" (Tabelle 21).

**Tabelle 21.** Differenz der Mittelwerte und Ergebnisse des Mittelwertvergleichs in der deutschen Experimentalgruppe (gepaarter t-Test) (n=22)

	Mittelwert	Std. Abweichung	Differenz der Mittelwerte	t	p	
FAHIGKEITEN	POST_ Interpretation	12.73	2.34	1.022	2,089	<b>.049</b>
	PRE_ Interpretation	11.71	2.68			
	POST_ Bewertung	8.44	1.50	.309	.978	.339
	PRE_ Bewertung	8.13	1.83			
	POST_ Analyse	12.16	2.96	1.359	2.286	<b>.033</b>
	PRE_ Analyse	10.80	3.26			
	POST_ Schlussfolgerung	12.05	2.93	-.182	-.305	.763
	PRE_ Schlussfolgerung	12.23	2.66			
	POST_ Erläuterung	12.29	2.98	.947	1.575	.130
	PRE_ Erläuterung	11.34	2.70			
	POST_ Selbstregulierung	8.47	1.88	.149	.349	.731
	PRE_ Selbstregulierung	8.32	1.87			
	POST_ Integrierter Wert	66.15	13.29	3.605	1.466	.157
	PRE_ Integrierter Wert	62.54	13.41			
DISPOSITIONS	POST_ Reflexion	5.53	.94	-.106	-.480	.636
	PRE_ Reflexion	5.64	.66			
	POST_ Achtsamkeit	4.33	1.30	.227	.814	.425
	PRE_ Achtsamkeit	4.10	1.38			
	POST_ Aufgeschlossenheit	4.91	1.27	-.215	-.631	.525
	PRE_ Aufgeschlossenheit	5.13	1.18			
	POST_ Organisation	4.95	1.42	.136	.483	.634
	PRE_ Organisation	4.82	1.28			

	Mittelwert	Std. Abweichung	Differenz der Mittelwerte	t	p
POST_ Durchhaltevermögen	5.18	1.11	-.273	-1.393	.178
PRE_ Durchhaltevermögen	5.45	.96			
POST_ Intrinsische Motivation	5.47	1.24	-.182	.774	.448
PRE_ Intrinsische Motivation	5.65	.89			
<b>DISPOSITIONEN</b>					
POST_ Integrierter Wert	30.37	4.68	-.413	-.479	.637
PRE_ Integrierter Wert	30.78	3.88			

Die pädagogischen Interventionen hatten keinen Einfluss auf die Ergebnisse der Dispositionen, möglicherweise aufgrund der geringen Anzahl der Stichprobe.

### Griechenland

Die griechische Experimentalstichprobe bestand aus 63 Studierenden, die im Fachbereich Pädagogik eingeschrieben waren; diese Population wies ein Durchschnittsalter von 24,48 Jahren auf (sd = 8,42; Bereich = 19 - 54). 63,5 % der Teilnehmer waren zwischen 20 und 21 Jahre alt. 85,7 % der Teilnehmer sind weiblich und 14,3 % männlich. Die Studierenden nahmen im Rahmen von drei Lehrveranstaltungen an dieser Studie teil: Didaktik der Naturwissenschaften (54%), Didaktik der biologischen Konzepte (34,9%) und Didaktik der Umweltkunde (11,1%). Tabelle 22 zeigt die deskriptiven Ergebnisse in jeder der Subskalen und Skalen für die experimentelle Gruppe zu Beginn der CTBACs.

**Tabelle 22.** Deskriptive Statistiken für CT-Fähigkeiten und CT-Dispositionen der griechischen Stichprobe zu Beginn der Pilotmodule.

	Dimensionen	Mittelwert	Std. Abweichung	Minimum	Maximum
Fähigkeiten	Interpretation	12.41	2.61	6.67	18.00
	Analyse	7.96	1.78	3.20	11.10
	Bewertung	11.97	2.51	5.20	17.50
	Schlussfolgerung	13.05	2.37	6.53	17.20
	Erläuterung	11.38	2.51	5.33	16.17
	Selbstregulierung	8.32	2.01	3.57	12.00

	Dimensionen	Mittelwert	Std. Abweichung	Minimum	Maximum
	<b>Integrierter Wert für CT-Fähigkeiten</b>	65.09	11.54	33.48	90.51
<b>Dispositionen</b>	<b>Reflektion</b>	5.88	.88	3.67	7.00
	<b>Achtsamkeit</b>	4.00	1.58	1.00	7.00
	<b>Aufgeschlossenheit</b>	5.06	1.20	2.00	7.00
	<b>Organisation</b>	4.90	1.45	1.00	7.00
	<b>Durchhaltevermögen</b>	5.49	1.11	2.67	7.00
	<b>Intrinsische Motivation</b>	5.24	1.20	1.75	7.00
	<b>Integrierter Wert für CT-Dispositionen</b>	30.56	4.22	20.83	38.75

Das Vorhandensein von Mittelwertunterschieden zwischen den Geschlechtern in der Phase vor der Intervention wurde aufgrund des enormen Ungleichgewichts zwischen den Größen beider Gruppen nicht bewertet. Der Vergleich zwischen den Modulen wurde nur zwischen den beiden Gruppen mit einer höheren Anzahl von Befragten durchgeführt: Didaktik der Naturwissenschaft (n = 34) und Didaktik der Biologie (n = 22) (Tabelle 23). Der Vergleich zeigt signifikante Unterschiede in den Ausgangswerten der Studierenden aus den beiden Modulen in zwei Fähigkeiten: *Interpretation* und *Analyse*. Am deutlichsten waren die Unterschiede jedoch in der Skala der CT-Dispositionen. In allen Fällen sind die Durchschnittswerte der Studierenden, die das Modul Didaktik der Naturwissenschaft besucht haben, höher als die der Studierenden, die das Modul Didaktik der Biologie besucht haben. Schließlich wurde ein positiver und signifikanter Zusammenhang zwischen dem Alter und der Disposition *Intrinsische Motivation* festgestellt ( $r=.337$ ;  $p=.007$ ).

**Tabelle 23.** Mittelwert der CT-Fähigkeiten und CT-Dispositionen beim Vortest bei griechischen Studierenden in den Modulen "Didaktik der Naturwissenschaft" (n=34) und "Didaktik der Biologie" (n=22).

	Vortest	Module	Mittelwert	Std. Abweichung	P
<b>Skills</b>	<b>Interpretation</b>	<b>Didaktik der Naturwissenschaft</b>	13.38	2.08	.007

Vortest	Module	Mittelwert	Std. Abweichung	P	
Bewertung	Didaktik der Biologie	11.38	2.81		
	Didaktik der Naturwissenschaft	8.13	1.60	.339	
	Didaktik der Biologie	7.60	2.17		
Analyse	Didaktik der Naturwissenschaft	12.51	2.07	.039	
	Didaktik der Biologie	11.07	3.02		
Schlussfolgerung	Didaktik der Naturwissenschaft	13.30	2.20	.218	
Erläuterung	Didaktik der Biologie	12.46	2.82		
	Didaktik der Naturwissenschaft	11.57	2.20	.498	
	Didaktik der Biologie	11.11	2.72		
Selfregulierung	Didaktik der Naturwissenschaft	8.62	1.99	.067	
	Didaktik der Biologie	7.58	2.03		
Fähigkeiten	Integrierter Wert	Didaktik der Naturwissenschaft	67.50	9.88	.053
		Didaktik der Biologie	61.20	13.91	
Reflektion	Didaktik der Naturwissenschaft	6.02	.87	.086	
	Didaktik der Biologie	5.61	.86		
	Didaktik der Naturwissenschaft	4.38	1.53	.024	
Achtsamkeit	Didaktik der Biologie	3.43	1.46		
	Didaktik der Naturwissenschaft	4.98	1.21	.946	
Aufgeschlossenheit	Didaktik der Biologie	5.00	1.18		
	Didaktik der Naturwissenschaft	5.31	1.32	.022	
	Didaktik der Biologie	4.41	1.50		
Dispositionen	Organisation	Didaktik der Naturwissenschaft	5.67	1.05	.086
		Didaktik der Biologie	5.15	1.11	
	Didaktik der Naturwissenschaft	5.65	1.09	.004	
Durchhaltevermögen	Didaktik der Biologie	4.68	1.20		
	Intrinsische Motivation	Didaktik der Naturwissenschaft			
	Didaktik der Biologie				

Vortest	Module	Mittelwert	Std. Abweichung	P
<b>Integrated Score</b>	<b>Didaktik der Naturwissenschaft</b>	32.01	4.05	.001
	<b>Didaktik der Biologie</b>	28.28	3.48	

Die allgemeinen Vor-Test/Nach-Test-Vergleiche (ohne Berücksichtigung der vorherigen Unterschiede) zeigen einen positiven Effekt der Intervention auf alle Fähigkeiten, wobei die Studierenden nach der Intervention höhere Mittelwerte aufweisen. Bei den CT-Dispositionen gibt es nur bei der *Aufgeschlossenheit* einen signifikanten Unterschied, allerdings in die entgegengesetzte Richtung als erwartet, mit einem höheren Mittelwert vor der Intervention als danach (Tabelle 24).

Im Allgemeinen wiesen die Studierenden des Moduls Didaktik der Naturwissenschaft nach den Interventionen höhere Zuwächse in den Bereichen der CT-Fähigkeiten und der integrierten Bewertung auf als die Studierenden des Moduls Didaktik der Biologie, mit Ausnahme der *Interpretation* (12,49 vs. 13,42). Auch bei den CT-Dispositionen waren die mittleren Unterschiede in den Dispositionswerten bei den Studierenden der Lehrveranstaltung Didaktik der Naturwissenschaft höher als bei den Studierenden der Lehrveranstaltung Didaktik der Biologie, mit Ausnahme der Dimension *Achtsamkeit* (3,44 vs. 3,96). Allerdings erreichten diese Unterschiede in den beiden Versuchsgruppen keine Signifikanz. Aufgrund der bereits bestehenden Unterschiede zwischen den Gruppen in den Kovariaten, nämlich für die Fähigkeiten "*Interpretation*" und "*Analyse*" sowie für 3 der 6 Dispositionen und den integrierten CT-Dispositionswert (wie oben dargestellt), sind die Ergebnisse jedoch mit Vorsicht zu interpretieren.

**Tabelle 24.** Differenz der Mittelwerte und Ergebnisse des Mittelwertvergleichs (gepaarter t-Test) in der griechischen Experimentalgruppe (n=63)

		Mittelwert	Std. Abweichung	Differenz der Mittelwerte	t	p
<b>Fähigkeiten</b>	<b>POST_ Interpretation</b>	13.26	2.54	.852	2.469	.016

	Mittelwert	Std. Abweichung	Differenz der Mittelwerte	t	p
PRE_ Interpretation	12.41	2.61			
POST_ Bewertung	8.86	1.99	.903	4.642	≤.0001
PRE_ Bewertung	7.96	1.78			
POST_ Analyse	13.09	2.91	1.123	4.642	≤.0001
PRE_ Analyse	11.97	2.51			
POST_ Schlussfolgerung	13.90	2.77	.853	4.642	≤.0001
PRE_ Schlussfolgerung	13.05	2.37			
POST_ Erläuterung	12.99	2.73	1.616	5.454	≤.0001
PRE_ Erläuterung	11.38	2.51			
POST_ Selbstregulierung	9.10	1.909	.782	3.890	≤.0001
PRE_ Selbstregulierung	8.32	2.01			
POST_ Integrierter Wert	71.22	13.23	6.130	5.412	≤.0001
PRE_ Integrierter Wert	65.09	11.54			
DISPOSITIONEN					
POST_ Reflektion	5.90	.90	.026	-.217	.829
PRE_ Reflektion	5.88	.88			
POST_ Achtsamkeit	3.84	1.69	-.159	.941	.350
PRE_ Achtsamkeit	4.00	1.58			
POST_ Aufgeschlossenheit	4.56	1.47	-.504	2.758	.008
PRE_ Aufgeschlossenheit	5.06	1.20			
POST_ Organisation	5.11	1.22	.206	-1.326	.190
PRE_ Organisation	4.90	1.45			
POST_ Durchhaltevermögen	5.56	1.05	.074	-.619	.538
PRE_ Durchhaltevermögen	5.49	1.11			
POST_ Intrinsische Motivation	5.30	1.20	.056	-.433	.667
PRE_ Intrinsische Motivation	5.24	1.20			
DISPOSITIONEN					
POST_ Integrierter Wert	30.56	4.22	.300	.709	.481
PRE_ Integrierter Wert	30.26	4.49			

Die Auswirkungen der pädagogischen Interventionen wurden auch zwischen den Versuchs- und Kontrollgruppen verglichen. Die Versuchsbedingung (Kontrolle vs. Experiment) wurde als unabhängige Variable, die Nachmessung als abhängige Variable und die Vormessung als Kovariate verwendet. Die Bestätigung des Nichtvorhandenseins früherer Unterschiede

zwischen den Gruppen in den Dimensionen Fähigkeiten und Dispositionen zeigte, dass die Annahme für alle Dimensionen erfüllt war, mit Ausnahme der Dimension *Aufgeschlossenheit*, in der die Experimentalgruppe einen höheren früheren Durchschnitt als die Kontrollgruppe aufwies (5,06 vs. 4,34). Die Versuchsgruppe zeigte höhere positive Veränderungen in der Bewertung der integrierten CT-Fähigkeiten ( $p=.007$ ) sowie in den Dimensionen "*Bewertung*" ( $p=0,040$ ), "*Schlussfolgerung*" ( $p=0,042$ ), "*Erläuterung*" ( $p=0,008$ ), aber nicht in "*Interpretation*", "*Analyse*" oder "*Selbstregulierung*". Ein ähnliches Muster wurde für den Zuwachs an CT-Dispositionen gefunden. Die Unterschiede waren in der Experimentalgruppe bei der integrierten Bewertung der Dispositionen ( $p=0,007$ ) und im Bereich "*Achtsamkeit*" ( $p=0,019$ ) höher, aber es wurden keine Unterschiede zwischen der Experimental- und der Kontrollgruppe bei den anderen Dispositionen festgestellt.

### **Litauen**

Die litauische Stichprobe besteht aus 20 Studenten des Studiengangs Internationale Beziehungen und Politikwissenschaft (BA), die im Modul Englisch als Fremdsprache eingeschrieben sind, mit einem Durchschnittsalter von 18,75 Jahren ( $sd = 0,639$ ; Spanne = 18 - 20). 65% der Teilnehmer waren weiblich und 35% waren männlich. Die Studierenden nahmen an dieser Studie im Rahmen des Pilotmoduls Englisch als Fremdsprache teil. Tabelle 25 zeigt die deskriptiven Ergebnisse für die einzelnen Dimensionen und Unterdimensionen des zur Datenerhebung verwendeten Instruments.

Die Überprüfung der vorherigen Gleichheit zwischen den Gruppen nach Alter oder Modul wurde aufgrund des Ungleichgewichts zwischen den beiden Gruppen nicht durchgeführt. Es wurde ein signifikanter, aber negativer Zusammenhang zwischen dem Alter und der CT-Disposition *Aufgeschlossenheit* festgestellt ( $r= -0,511$ ;  $p = 0,021$ ).

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

	Dimensionen	Mittelwert	Std. Abweichung	Minimum	Maximum
<b>Fähigkeiten</b>	Interpretation	11.47	2.20	6.83	15.50
	Analyse	7.98	1.91	3.90	11.50
	Bewertung	9.80	3.43	4.50	17.25
	Schlussfolgerung	10.89	3.25	4.53	17.40
	Erläuterung	10.95	3.09	6.50	16.67
	Selbstregulierung	7.32	2.28	3.00	11.19
	Integrierter Wert für CT-Fähigkeiten	58.40	13.83	38.13	88.01
<b>Dispositionen</b>	Reflektion	5.57	.77	4.00	7.00
	Achtsamkeit	4.08	1.41	1.50	6.75
	Aufgeschlossenheit	5.51	.80	4.25	6.75
	Organisation	4.22	1.66	1.00	6.67
	Durchhaltevermögen	5.00	1.14	3.00	7.00
	Intrinsische Motivation	5.14	1.16	2.50	7.00
	Integrierter Wert für CT-Dispositionen	29.51	4.47	20.58	38.75

Die Vergleiche zwischen Vor- und Nachtests ergaben einen positiven Effekt der Intervention bei allen bis auf eine CT-Fähigkeit: "Bewertung" (Tabelle 26). Im Falle Litauens wurde nur ein einziges Modul betrachtet und daher wurden keine Unterschiede in den Lernstrategien berücksichtigt.

**Tabelle 26.** Differenz der Mittelwerte und Ergebnisse des Vergleichs der Mittelwerte (gepaarter t-Test) bei den litauischen Studierenden (n=20)

	Mean	Std. Deviation	Mean difference	t	p	
<b>FÄHIGKEITEN</b>	POST_ Interpretation	13.43	2.14	1.958	-3.738	.001
	PRE_ Interpretation	11.47	2.20			
	POST_ Bewertung	8.63	1.94	.655	-1.489	.153
	PRE_ Bewertung	7.98	1.91			
	POST_ Analyse	12.63	2.80	2.835	-4.281	≤.0001
	PRE_ Analyse	9.80	3.43			
	POST_ Schlussfolgerung	12.60	2.74	1.717	-3.321	.004

		Mean	Std. Deviation	Mean difference	t	p
	<b>PRE_ Schlussfolgerung</b>	10.89	3.25			
<b>FAHIGKEITEN</b>	<b>POST_ Erläuterung</b>	13.08	2.50	2.125	-4.959	≤.0001
	<b>PRE_ Erläuterung</b>	10.95	3.09			
	<b>POST_ Selbstregulierung</b>	8.58	2.07	1.257	-3.110	.006
	<b>PRE_ Selbstregulierung</b>	7.32	2.28			
	<b>POST_ Integrierter Wert</b>	68.95	12.63	10.547	4.644	≤.0001
	<b>PRE_ Integrierter Wert</b>	58.40	13.83			
<b>DISPOSITIONEN</b>	<b>POST_ Reflektion</b>	5.77	.88	.200	1.092	.289
	<b>PRE_ Reflektion</b>	5.57	.77			
	<b>POST_ Achtsamkeit</b>	4.31	1.61	.238	.937	.361
	<b>PRE_ Achtsamkeit</b>	4.08	1.41			
	<b>POST_ Aufgeschlossenheit</b>	5.64	.58	.125	1.022	.320
	<b>PRE_ Aufgeschlossenheit</b>	5.51	.80			
	<b>POST_ Organisation</b>	4.48	1.73	.267	1.823	.084
	<b>PRE_ Organisation</b>	4.22	1.66			
	<b>POST_ Durchhaltevermögen</b>	4.67	1.59	-.333	-1.541	.140
	<b>PRE_ Durchhaltevermögen</b>	5.00	1.14			
	<b>POST_ Intrinsische Motivation</b>	5.08	1.44	-.063	-.260	.798
	<b>PRE_ Intrinsische Motivation</b>	5.14	1.16			
	<b>POST_ Integrierter Wert</b>	29.94	5.73	.433	.712	.485
	<b>PRE_ Integrierter Wert</b>	29.51	4.47			

### Portugal

Die portugiesische Versuchsgruppe bestand aus 100 Studierenden, die im Fach Veterinärmedizin (Integrierter Master) eingeschrieben waren, mit einem Durchschnittsalter von 22,32 Jahren (sd = 4,62; Bereich = 19 - 52). 79% der Teilnehmer waren weiblich und 21% waren männlich. Die Studierenden nahmen an dieser Studie im Rahmen von drei Pilotmodule teil: Bildgebung (36 %), Deontologie (42 %) und Gynäkologie, Andrologie und Geburtshilfe (22 %). Tabelle 27 zeigt die deskriptiven Ergebnisse für jede Dimension und Subdimension der Skalen CT-Fähigkeiten und CT-Disposition.

Der Vergleich der Mittelwerte zwischen den Geschlechtern zu Beginn des Semesters zeigt, dass es keine signifikanten Unterschiede zwischen männlichen und weiblichen Studenten auf den Skalen der CT-Fähigkeiten gibt, außer bei der Fähigkeit "*Analyse*" ( $p = .035$ ). In allen Fällen wiesen die Männer höhere Mittelwerte auf als die Frauen.

**Tabelle 27.** Deskriptive Statistiken für CT-Fähigkeiten und CT-Dispositionen der portugiesischen Stichprobe zu Beginn der Pilotmodule

	Vortest Werte	Mittelwert	Std. Abweichung	Minimum	Maximum
<b>Fähigkeiten</b>	<b>Interpretation</b>	12.76	1.87	7,50	16,75
	<b>Analyse</b>	7.62	1.73	2,60	11,60
	<b>Bewertung</b>	11.58	2.14	5,75	15,75
	<b>Schlussfolgerung</b>	11.72	2.34	4,93	16,40
	<b>Erläuterung</b>	11.04	2.42	5,17	16,17
	<b>Selbstregulierung</b>	8.56	1.71	5,14	11,71
	<b>Integrierter Wert für CT-Fähigkeiten</b>	63.28	10.06	37,91	86,55
<b>Dispositionen</b>	<b>Reflektion</b>	5.57	1.01	1,67	7,00
	<b>Achtsamkeit</b>	3.87	1.40	1,00	7,00
	<b>Aufgeschlossenheit</b>	5.91	1.08	1,00	7,00
	<b>Organisation</b>	4.95	1.49	1,00	7,00
	<b>Durchhaltevermögen</b>	5.31	1.18	1,67	7,00
	<b>Intrinsische Motivation</b>	5.32	1.12	1,50	7,00
	<b>Integrierter Wert für CT-Dispositionen</b>	30.93	4.60	19,50	41,50

Bei den CT-Dispositionen wurden signifikante Unterschiede bei der "*Organisation*" festgestellt ( $p = 0,001$ ), wobei Frauen einen höheren Durchschnitt als Männer aufwiesen. Das Alter zeigte keinen signifikanten Zusammenhang mit einer der Dimensionen der CT-Fähigkeiten oder CT-Dispositionen.

Auf der Skala der CT-Fähigkeiten zeigte der Vergleich zwischen den Modulen signifikante Unterschiede in den Mittelwerten vor den Interventionen, insbesondere bei den Fähigkeiten "*Analyse*" ( $p = .034$ ) und "*Erläuterung*" ( $p = .042$ ). In allen Fällen hatten die Studierenden der

Gynäkologie höhere Durchschnittswerte als die Studierenden der Bildgebung, die ihrerseits vor der Intervention höhere Durchschnittswerte als die Studierenden der Deontologie hatten. Hinsichtlich der Analyse- und Erklärungsfähigkeiten ergaben die Bonferroni-Tests signifikante Unterschiede nur zwischen den Gynäkologie- und Deontologiemodulen ( $p = 0,040$  bzw.  $p = 0,041$ ), wobei die Studierenden der Gynäkologie einen höheren Durchschnitt ( $M = 12,61$  bzw.  $M = 11,97$ ) als die Studierenden der Deontologie ( $M = 11,23$  bzw.  $M = 10,41$ ) aufwiesen. Es gibt keine Unterschiede zwischen diesen beiden Gruppen und den Imaging-Studenten, die einen Durchschnitt von 11,35 bzw. 11,20 aufweisen. Im Gegensatz dazu gibt es keine signifikanten Unterschiede in den CT-Dispositionen vor der Intervention, die mit dem Modul in Verbindung stehen könnten.

Die Gesamtveränderungen als Reaktion auf die pädagogischen Interventionen (ohne Berücksichtigung der zuvor festgestellten Unterschiede bei einigen Fähigkeiten in Abhängigkeit von Geschlecht und Modul) zeigen einen positiven Effekt der Intervention mit signifikant höheren Mittelwerten bei allen Fähigkeitsdimensionen. Im Falle der CT-Dispositionen sind die Gewinne durch die Intervention jedoch nicht so eindeutig. Trotz des signifikanten Gesamtgewinns ( $p = 0,010$ ) zeigten die Studierenden nur in drei der sechs Dispositionen signifikant höhere Werte, und in zwei von ihnen verschlechterten sich die Ergebnisse nach der Intervention, obwohl die Unterschiede keine statistische Signifikanz erreichten (Tabelle 28).

Die differentielle Wirkung der in den drei Modulen durchgeführten Interventionen wurde analysiert, aber es wurden keine differentiellen Auswirkungen für das Modul auf die von den Studierenden erzielten Ergebnisse gefunden (Tabelle 29).

**Tabelle 28.** Differenz der Mittelwerte und Ergebnisse des Vergleichs der Mittelwerte (gepaarter t-Test) in der portugiesischen Experimentalgruppe (n=100)

		Mittelwert	Std. Abweichung	Differenz der Mittelwerte	t	p
<b>FÄHIGKEITEN</b>	POST_ Interpretation	13.85	1.90	1.092	5.028	≤.0001
	PRE_ Interpretation	12.76	1.87			
	POST_ Bewertung	8.70	1.54	1.076	5.983	≤.0001
	PRE_ Bewertung	7.62	1.73			
	POST_ Analyse	12.98	2.14	1.405	6.121	≤.0001
	PRE_ Analyse	11.58	2.14			
	POST_ Schlussfolgerung	13.24	2.03	1.521	6.648	≤.0001
	PRE_ Schlussfolgerung	11.72	2.34			
	POST_ Erläuterung	12.46	2.22	1.420	5.919	≤.0001
	PRE_ Erläuterung	11.04	2.42			
	POST_ Selbstregulierung	9.23	1.50	.677	4.274	≤.0001
	PRE_ Selbstregulierung	8.56	1.71			
	POST_ Integrierter Wert	70.47	9.98	7.190	7.211	≤.0001
	PRE_ Integrierter Wert	63.28	10.06			
<b>DISPOSITIONEN</b>	POST_ Reflektion	5.90	.81690	.330	3.590	.001
	PRE_ Reflektion	5.57	1.01			
	POST_ Achtsamkeit	3.76	1.43	-.113	-.966	.336
	PRE_ Achtsamkeit	3.87	1.40			
	POST_ Aufgeschlossenheit	5.913	1.30	.005	.042	.966
	PRE_ Aufgeschlossenheit	5.907	1.08			
	POST_ Organisation	5.09	1.38	.137	1.215	.227
	PRE_ Organisation	4.95	1.49			
	POST_ Durchhaltevermögen	5.55	1.18	.247	2.899	.005
	PRE_ Durchhaltevermögen	5.31	1.18			
	POST_ Intrinsische Motivation	5.64	.95	.320	3.655	≤.0001
	PRE_ Intrinsische Motivation	5.32	1.12			
	POST_ Integrierter Wert	31.85	4.54	.916	2.618	.010
	PRE_ Integrierter Wert	30.93	4.60			



**Tabelle 29:** Modulbezogene Auswirkungen auf den durchschnittlichen Zuwachs ( $\pm$ sd) an CT-Fähigkeiten und CT-Dispositionen in der portugiesischen Experimentalgruppe vor der Einführung der CTBACs

	Dimensionen	Bildgebung (n=36)	Deontologie (n=42)	Gynäkologie (n=22)	F (df = 2)	p	Eta
Fähigkeiten	Interpretation	13,37 $\pm$ 2.34	14.25 $\pm$ 1.67	13.88 $\pm$ 1.31	2.079	0.131	0.042
	Analyse	12,6 $\pm$ 2.53	13.19 $\pm$ 1.93	13.16 $\pm$ 1.81	1.224	0.299	0.025
	Bewertung	8.41 $\pm$ 1.54	8.90 $\pm$ 1.47	8.78 $\pm$ 1.38	1.952	0.148	0.039
	Schlussfolgerung	12.84 $\pm$ 2.28	13.35 $\pm$ 1.94	13.69 $\pm$ 1.75	1.356	0.263	0.027
	Erläuterung	12.05 $\pm$ 2.37	12.43 $\pm$ 2.11	13.17 $\pm$ 2.07	1.669	0.194	0.034
	Selbstregulierung	9.25 $\pm$ 1.71	9.18 $\pm$ 1.37	9.32 $\pm$ 1.46	0.013	0.987	0.000
	Integrierter Wert für CT-Fähigkeiten	68.55 $\pm$ 11.65	71.30 $\pm$ 9.08	72.00 $\pm$ 8.52	1.568	0.214	0.032
Dispositionen	Reflektion	5.94 $\pm$ 0.90	5.87 $\pm$ 0.81	5.92 $\pm$ 0.72	0.279	0.757	0.006
	Achtsamkeit	4.13 $\pm$ 1.36	3.63 $\pm$ 1.43	3.41 $\pm$ 1.45	1.680	0.192	0.034
	Aufgeschlossenheit	6.21 $\pm$ 0.94	5.64 $\pm$ 1.44	5.92 $\pm$ 1.44	1.821	0.167	0.037
	Organisation	5.31 $\pm$ 1.35	4.91 $\pm$ 1.34	5.06 $\pm$ 1.51	0.278	0.758	0.006
	Durchhaltevermögen	5.78 $\pm$ 1.23	5.33 $\pm$ 1.22	5.62 $\pm$ 0.93	1.174	0.313	0.024
	Intrinsische Motivation	5.80 $\pm$ 1.01	5.48 $\pm$ 0.90	5.68 $\pm$ 0.96	0.613	0.544	0.013
	Integrierter Wert für CT-Dispositionen	33.15 $\pm$ 4.74	30.85 $\pm$ 4.40	31.62 $\pm$ 4.12	0.614	0.543	0.013

## Rumänien

Die rumänische Experimentalgruppe setzte sich aus 81 Studierenden des Fachbereichs Betriebs- und Volkswirtschaft zusammen, mit einem Durchschnittsalter von 26,58 Jahren (sd = 10,04; Rang = 19-52). 65 % der Teilnehmer sind zwischen 19 und 23 Jahre alt. 84% der Teilnehmer sind weiblich und 16% männlich. Die Studenten nahmen an dieser Studie im Rahmen von drei Kursen teil: Wirtschaftskommunikation (38,3%), Pädagogik und Didaktik der Finanzbuchhaltung (39,5%) und Virtuelle Lernumgebungen in der Wirtschaftswissenschaft (22,2%). Tabelle 30 zeigt die deskriptiven Ergebnisse in jeder der Dimensionen der CT-Fähigkeiten und CT-Dispositionen.

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

	Dimensionen	Mittelwert	Std. Abweichung	Minimum	Maximum
Fähigkeiten	Interpretation	13.28	2.29	8.25	18.00
	Analyse	8.92	1.81	3.90	11.80
	Bewertung	13.26	2.68	6.75	17.40
	Schlussfolgerung	13.40	2.59	6.87	17.60
	Erläuterung	12.20	2.84	6.00	18.00
	Selbstregulierung	9.15	1.91	5.10	12.00
	Integrierter Wert für CT-Fähigkeiten	70.20	12.26	42.41	92.65
Dispositionen	Reflektion	5.91	.80	4.00	7.00
	Achtsamkeit	4.11	1.34	1.75	7.00
	Aufgeschlossenheit	5.18	1.28	2.25	7.00
	Organisation	5.10	1.22	2.33	7.00
	Durchhaltevermögen	5.54	1.08	2.33	7.00
	Intrinsische Motivation	5.59	.91	3.00	7.00
	Integrierter Wert für CT-Dispositionen	31.44	4.15	19.50	39.75

Der Vergleich der Mittelwerte zwischen den Geschlechtern in der Phase vor der Intervention zeigt, dass es keine signifikanten Unterschiede zwischen männlichen und weiblichen

Studierenden gibt, weder in den Dimensionen der CT-Fähigkeiten noch in den CT-Dispositionen, mit Ausnahme der Fähigkeit "Analyse" (11,76±2,82 vs. 13,55±2,57) und der Disposition "Durchhaltevermögen" (4,97±1,23 vs. 5,65±1,03).

Der Vergleich zwischen den drei Pilotmodulen zeigt signifikante Unterschiede in den Ausgangswerten der Studierenden aus den verschiedenen Modulen in der Fähigkeit "Interpretation" und in der Disposition "Durchhaltevermögen". In beiden Fällen bestätigt der Bonferroni-Post-hoc-Test, dass die Unterschiede zwischen den Modulen "Pädagogik und Didaktik der Finanzbuchhaltung" und "Virtuelle Lernumgebungen in Wirtschaftswissenschaften" bestehen. Ebenfalls in beiden Fällen weisen die Studierenden der Virtuellen Lernumgebungen eine höhere durchschnittliche Punktzahl in den Bereichen Interpretation (14,35 vs. 12,67) und Ausdauer (6,00 vs. 5,20) auf als die Studierenden der Pädagogik und Didaktik der Finanzbuchhaltung.

Was schließlich das Alter betrifft, werden zwei positive und signifikante Zusammenhänge beobachtet, und zwar bei denselben Fähigkeiten und Dispositionen wie zuvor erwähnt: *Interpretation* ( $p=.004$ ) und *Durchhaltevermögen* ( $p=.020$ ).

Insgesamt zeigen die Vergleiche zwischen Vor- und Nachtest (ohne Berücksichtigung der vorherigen Unterschiede) einen positiven Effekt der Intervention auf nur zwei Fähigkeiten (*Erläuterung* und *Interpretation*) und auf die integrierte Bewertung der CT-Fähigkeiten (Tabelle 31).

**Tabelle 31.** Differenz der Mittelwerte und Ergebnisse des Vergleichs der Mittelwerte (gepaarter t-Test) bei den rumänischen Studierenden (n=81)

		Mittelwert	Std. Abweichung	Differenz der Mittelwerte	t	p
FAHIGKEITEN	POST_ Interpretation	14.09	2.16	.809	3.346	.001
	PRE_ Interpretation	13.28	2.29			
	POST_ Bewertung	9.21	1.68	.295	1.571	.120
	PRE_ Bewertung	8.92	1.81			

	POST_ Analyse	13.55	2.39	.290	1.043	.300
	PRE_ Analyse	13.26	2.68			
	POST_ Schlussfolgerung	13.85	2.49	.456	1.727	.088
	PRE_ Schlussfolgerung	13.40	2.59			
	POST_ Erläuterung	13.43	2.43	1.237	4.092	≤.0001
	PRE_ Erläuterung	12.20	2.84	1.237	4.092	≤.0001
	POST_ Selbstregulierung	9.49	1.82	.346	1.886	.063
	PRE_ Selbstregulierung	9.15	1.91			
	POST_ Integrierter Wert	73.63	11.67	3.431	2.944	.004
	PRE_ Integrierter Wert	70.20	12.26			
DISPOSITIONEN	POST_ Reflektion	5.81	.93	-.103	-1.038	.303
	PRE_ Reflektion	5.91	.80			
	POST_ Achtsamkeit	4.04	1.31	-.065	-.474	.637
	PRE_ Achtsamkeit	4.11	1.34			
	POST_ Aufgeschlossenheit	4.89	1.38	-.290	-1.969	.052
	PRE_ Aufgeschlossenheit	5.18	1.28			
	POST_ Organisation	5.30	1.15	.202	1.407	.163
	PRE_ Organisation	5.10	1.22			
	POST_ Durchhaltevermögen	5.58	1.08	.037	.293	.771
	PRE_ Durchhaltevermögen	5.54	1.08			
	POST_ Intrinsische Motivation	5.56	.98	-.031	-.307	.760
	PRE_ Intrinsische Motivation	5.59	.91			
	POST_ Integrierter Wert	31.19	4.53	-.250	-.580	.564
	PRE_ Integrierter Wert	31.44	4.15			

Der GLM- Univariate ANCOVA-Test wurde berechnet, um die Zuwächse in den Pilotmodulen [Wirtschaftskommunikation (n=31), Pädagogik und Didaktik der Finanzbuchhaltung (n=32) und Virtuelle Lernumgebungen in Wirtschaftswissenschaften (n= 18)] zu schätzen. Dennoch ist bei der Interpretation der ermittelten Unterschiede Vorsicht geboten, da die Annahme nicht vorhandener vorheriger Unterschiede zwischen den Gruppen in der Kovariate für die Fähigkeit *Interpretation* und die Disposition *Durchhaltevermögen* nicht erfüllt war (siehe vorherige Absätze).

Im Allgemeinen unterschieden sich die interventionsassoziierten Zuwächse bei den CT-Fähigkeiten nicht zwischen den drei Pilotmodulen, mit Ausnahme der integrierten Punktzahl für die CT-Dispositionen, die bei den Studierenden der Wirtschaftskommunikation und der Virtuellen Lernumgebungen in Wirtschaftswissenschaften ( $32,45 \pm 4,613$  bzw.  $32,25 \pm 3,78$ ) höher war ( $p=0,017$ ) als bei den Studierenden des Moduls Pädagogik und Didaktik der Finanzbuchhaltung ( $29,38 \pm 4,34$ ). Was den Zuwachs an Dispositionen betrifft, so wurden die Unterschiede zwischen den Modulen in folgenden Bereichen festgestellt: "*Achtsamkeit*" ( $p=.028$ ), wobei die Studierenden des Moduls Wirtschaftskommunikation einen höheren Zuwachs verzeichneten als die Studierenden des Moduls Virtuelle Lernumgebungen in Wirtschaftswissenschaften oder Pädagogik und Didaktik der Finanzbuchhaltung ( $4,45 \pm 1,31$  vs.  $3,73 \pm 1,21$  vs.  $3,89 \pm 1,38$ ); "*Aufgeschlossenheit*" ( $p=0,047$ ) zeigen die Studierenden in den Modulen Wirtschaftskommunikation und Pädagogik und Didaktik der Finanzbuchhaltung höhere Werte als die Studierenden in den Virtuellen Lernumgebungen in Wirtschaftswissenschaften ( $5,14 \pm 1,40$  und  $5,22 \pm 1,30$  vs.  $4,47 \pm 1,33$  vs.); und in "*Intrinsische Motivation*" ( $p=.009$ ). In der letztgenannten Kategorie zeigten sowohl die Studierenden in den Modulen Wirtschaftskommunikation und -pädagogik als auch Didaktik der Finanzbuchhaltung höhere Zuwächse ( $5,92 \pm 1,02$  und  $5,75 \pm 0,76$ ) als die Studierenden in den virtuellen Lernumgebungen der Wirtschaftswissenschaften ( $5,11 \pm 0,91$ ).

### 3. Diskussion

An der CTBAC-Pilotphase nahmen 609 Studierende aus den fünf Partnerländern teil, die in verschiedenen Kursen innerhalb eines bestimmten Studiengangs oder einer bestimmten Disziplin eingeschrieben waren (siehe Tabelle 1). Diese Zahl ist wesentlich höher als die ursprünglich im Antragsformular des Think4Jobs-Projekts vorgeschlagene Zahl (damals sollten 150 Studierende in CTBAC-Kursen eingeschrieben werden) und ergibt sich zumindest teilweise aus der Uneinheitlichkeit der Studierenden, die innerhalb derselben Einrichtung und desselben akademischen Jahres nach unterschiedlichen Strategien oder Ansätzen lernen. Daher war es in einigen Ländern, in denen derselbe Kurs nicht in zwei aufeinanderfolgenden

Semestern für eine unterschiedliche Gruppe von Studierenden angeboten wurde, nicht möglich, genügend Studierende für die Bildung einer Kontrollgruppe zu gewinnen. Es ist möglich, dass die Studierenden die vorgeschlagenen Unterschiede im Lernprozess entweder als unfair oder nachteilig empfanden - da sie zu unterschiedlichen Lernniveaus führten -, oder dass sie möglicherweise eine Situation herbeiführten, die ihre Arbeitsbelastung erhöhte, ohne ein gleichwertiges Lernen zu garantieren, und den damit verbundenen Gewinn nicht wahrnahmen.

Kontrollgruppen wurden nur in Griechenland (40 Studierende, die im Rahmen des Lehrerausbildungsprogramms in den Kurs "Didaktik der Naturwissenschaft" eingeschrieben waren) und in Portugal für die Lehrplan-Praktika des Veterinärmedizin-Programms (n=9), die am Hospital Veterinário do Atlântico entwickelt wurden, gebildet.

Von den Studierenden, die an der CTBAC-Pilotstudie teilnahmen, beantworteten nur 54 % (n=286) die Fragebögen vor und nach dem Test, mit denen die CT-Fähigkeiten und CT-Dispositionen der Studierenden vor und nach den Pilotmaßnahmen bewertet werden sollten. Dennoch ist diese Zahl ausreichend, um die pädagogischen Interventionen zu unterstützen und zu validieren. In der Anwendung schlugen wir vor, die Entwicklung der CT-Fähigkeiten und CT-Dispositionen zu drei verschiedenen Zeitpunkten des CTBAC-Pilotprojekts zu bewerten. Obwohl die Fragebögen, die zur Bewertung der CT bei den Studierenden verwendet wurden, zu den drei vorgeschlagenen Zeitpunkten (Vortest, Zwischentest und Nachtest) ausgefüllt wurden, nahm die Anzahl der Studierenden, die geantwortet haben, vom Vortest bis zur Messung nach dem Test ab, insbesondere in den Modulen, die im ersten Semester durchgeführt wurden. Um Informationsverluste zu vermeiden, wurde daher beschlossen, nur die Antworten auf die erste und die letzte Erhebung des ersten Semesters zu verwenden und den Fragebogen im zweiten Semester zweimal zu wiederholen.

Obwohl die Aufforderung zum Ausfüllen der Fragebögen über Moodle der CTBACs an die Studierenden geschickt wurde und die Plattform eine E-Mail-Benachrichtigung für die

Aufgaben und Aktivitäten bereitstellt, wurde festgestellt, dass dies einen geringen Einfluss auf die Befolgung der Aufforderung durch die Befragten hat [44]. Da die Nichterfüllung der Aufgaben (Beantwortung der Fragebögen) keine Konsequenzen nach sich zog und die Befragten auch nicht mit Geld (oder Gutscheinen oder anderen Vorteilen) belohnt wurden, fühlten sich die Studierenden nicht verpflichtet, die Fragebögen auszufüllen. Über dieses Problem, Studierende zur Beantwortung einer Vielzahl von Feedback-Fragebögen zu bewegen, wurde bereits früher berichtet [24, 45]. Es wurde vermutet, dass die schwachen Rücklaufquoten bei Fragebögen im Hochschulkontext damit zusammenhängen, dass die Studierenden nicht erwarten, dass die aus den Fragebögen gewonnenen Daten in einem Bereich, der für ihre persönlichen Ziele von Interesse sein könnte, sinnvoll genutzt werden. In Ermangelung eines verstärkenden Reizes, der ihr Engagement auslöst, was zu einer Verzerrung der Konditionierung führen könnte [44], bleibt die Zahl der Befragten daher tendenziell niedrig und verringert sich jedes Mal, wenn derselbe Fragebogen an sie weitergegeben wird. Die Anzahl der Fragen wurde reduziert, um die Belastung durch die Länge zu minimieren, die Bedeutung und die Ziele der Fragebögen (Validierung der Wirksamkeit der Unterrichtsstrategien) wurden betont und die Hauptkonzepte (CT-Fähigkeiten und CT-Dispositionen und wie sie von Fachleuten verwendet werden) wurden zu Beginn des Kurses erklärt, zusammen mit Empfehlungen zum Ausfüllen der Google-Formulare, in dem Versuch, die Studenten zum Beantworten der Fragebögen zu bewegen. Wir kennen die genauen Gründe für die niedrige Rücklaufquote des Fragebogens nicht, aber wir glauben, dass die Befragten ehrlich und bereit waren, einen unvoreingenommenen Bericht über ihre wahrgenommenen CT-Fähigkeiten und CT-Dispositionen abzugeben.

Die Anzahl der Befragten war ungleichmäßig auf die einzelnen Länder verteilt, aber die Fragebögen wurden freiwillig ausgefüllt, ohne Zwang oder Konsequenzen für Studierende, die nicht geantwortet haben. Dennoch reichte die Gesamtzahl der gepaarten Fragebögen aus, um eine zuverlässige globale statistische Analyse durchzuführen und die Ergebnisse der CT-Fähigkeiten und CT-Dispositionen zwischen den drei am stärksten vertretenen Ländern (Griechenland, Portugal und Rumänien) zu vergleichen.

Die Baseline-Analyse belegte die Existenz früherer Unterschiede im Durchschnitt und in den Werten der integrierten Skala der CT-Disposition in Abhängigkeit vom Geschlecht. Auch bei der integrierten Bewertung der CT-Fähigkeiten und den Kompetenzdimensionen wurden Unterschiede zwischen den Ländern festgestellt; diese Unterschiede beschränkten sich auf die Dimension *Aufgeschlossenheit*.

Das Alter stand in einem positiven Zusammenhang mit den Werten für die CT-Fähigkeiten, im Falle der CT-Dispositionen jedoch nur für den integrierten Wert und drei der sechs Dimensionen. Ein Alterseffekt bei den CT-Fähigkeiten wurde bereits früher berichtet [46]. Wir können jedoch die Hypothese nicht ausschließen, dass der Alterseffekt andere externe Faktoren verdeckt, wie z. B. die Auswirkungen des Landes oder der Disziplin. Oft ist der Alterseffekt bei CT weniger sichtbar, wenn wir ein kurzes Altersintervall innerhalb der Population haben, und ausgeprägter, wenn wir verschiedene Reifegrade über das Alter hinweg vergleichen [47] und Studierende in verschiedenen Stufen eines Studiengangs [48]. Die Altersverteilung der Studenten, die die Experimental-/Interventionsgruppe bildeten, war schief, aber trotz des höheren Alters der Studenten waren die meisten Studenten zwischen 19 und 21 Jahre alt (75,6 %), 18,2 % der Studenten waren zwischen 22 und 28 Jahre alt und nur knapp 6 % von ihnen waren über 38 Jahre alt.

Die festgestellten Ländereffekte überschneiden sich mit möglichen Effekten, die sich aus den verschiedenen Bereichen der überwachten Programme ergeben, da jedes Land eine andere Disziplin vertritt. Aufgrund des Projektdesigns kann dieser Zusammenhang nicht unterschieden werden.

Die durch die Baseline-Analyse aufgedeckten Unterschiede raten dazu, beim Vergleich der Ergebnisse zwischen den Geschlechtern und den Ländern die bestehenden Unterschiede vor und nach dem Test zu berücksichtigen. Die Untersuchung der Auswirkungen der Interventionen sollte für jedes Geschlecht und jedes Land getrennt bewertet werden, um die tatsächliche Wirksamkeit einer Intervention zu beurteilen. Aufgrund des unausgewogenen

Geschlechterverhältnisses in der Bevölkerung und der Repräsentativität der Länder wurde dieser Vergleich für diesen Bericht jedoch nicht durchgeführt.

### ***CTBAC-assozierte Verbesserungen der CT-Fähigkeiten und CT-Dispositionen***

Der Vergleich der Ergebnisse vor und nach dem Test zeigte, dass die in den Pilot-CTBACs durchgeführten Interventionen den Studierenden zwar eine Verbesserung ihrer CT-Fähigkeiten ermöglichten, die Auswirkungen auf die CT-Dispositionen jedoch geringer waren und oft keine Signifikanz erreichten. Dies unterstützt die Idee, dass es schwieriger ist, Dispositionen über kurze Zeiträume (wie ein Semester) zu verändern, da dies eine kontinuierliche und konzentrierte Praxis und eine kontinuierliche Reflexion erfordert, damit sie allmählich verinnerlicht werden können. Da es sich bei den Dispositionen um Einstellungsmerkmale handelt, erfordern sie außerdem eine intrinsische Bereitschaft und Anstrengung, um sie zu kultivieren, während die CT-Fähigkeiten ein prozedurales Verhalten darstellen, das als Reaktion auf einen Auslöser (Training) aufgenommen werden kann. Außerdem lassen sich verschiedene Dispositionen möglicherweise leichter fördern als andere, was erklären könnte, warum die Interventionen bei einigen Dispositionen zu Verbesserungen führten, bei anderen jedoch nicht.

Um die Auswirkungen der CTBACs auf die CT-Entwicklung besser einschätzen zu können, wurden die Fortschritte bei den Fähigkeiten und Neigungen zwischen den drei wichtigsten repräsentativen Ländern und innerhalb jedes Landes verglichen, da sich auch die umgesetzten Strategien und die mit den pädagogischen Interventionen angestrebten CT-Fähigkeiten und Neigungen zwischen den Ländern/Programmen unterscheiden.

Der durchschnittliche Zuwachs bei den integrierten CT-Fähigkeiten und -Dispositionen war bei den portugiesischen Studierenden höher, gefolgt von den griechischen Studierenden. Diese Beobachtung könnte darauf zurückzuführen sein, dass die rumänischen Studierenden die CTBACs im Allgemeinen mit einer höheren Punktzahl abgeschlossen haben, oder sie könnte auch auf Unterschiede in den pädagogischen Ansätzen zurückzuführen sein, die bei den

Interventionen verwendet wurden. Die in der Veterinärmedizin durchgeführten Aktivitäten hatten beispielsweise einen engeren Rahmen, was zu den etwas höheren Ergebnissen beitragen könnte.

## Deutschland

Die Hochschule Emden/Leer (HSEL) und die Orgadata AG (Orgadata) haben im Studienjahr 2021-2022 vier Critical Thinking Apprenticeships Curricula (CTAC) im Fach Wirtschaftsinformatik implementiert. Im Einzelnen wurden zwei Module "Design Patterns" (n=8) und "Economic Aspects of Industrial Digitalization" (n=8) im Wintersemester 2021-2022 und die beiden anderen Module "Innovationsmanagement" und "Wissenschaftliches Seminar" im Sommersemester 2022 implementiert. Das Modul "Design Patterns" war ein Pflichtkurs, der für die Auszubildenden von Orgadata angeboten wurde, während die anderen Module Wahlmodule waren, die für die Masterstudenten der HSEL angeboten wurden. Die ersten beiden Kurse wurden wie in Mäkiö und Kollegen [1] beschrieben durchgeführt und über einen Zeitraum von 16 Wochen, 1,5 Stunden pro Woche, im Klassenverband unterrichtet. Die beiden anderen Module wurden wie in Mäkiö & Mäkiö [49] beschrieben durchgeführt. Die Studierenden, die diese Module besuchten, bildeten die Experimentalgruppe der geplanten Interventionen. Da nur eine kleine Anzahl von Studierenden an den Modulen und den Erhebungen zur Selbsteinschätzung teilnahm, war die Stichprobengröße für statistische Schlussfolgerungen zu gering.

Aufgrund ihrer beruflichen Erfahrung sowohl in der Wirtschaft als auch im Bildungswesen waren sich die HSEL- und Orgadata-Lehrkräfte der Notwendigkeit bewusst, ihren Studierenden CT-Fähigkeiten und CT-Dispositionen zu vermitteln. Daher *haben* sie in diesen Kursen CT-Fähigkeiten wie *Interpretation, Analyse, Bewertung* und Erläuterung behandelt. Vergleiche zwischen Vor- und Nachtests deuten darauf hin, dass sich die Interventionen insgesamt positiv auf die Entwicklung der Fähigkeiten *Interpretation* und *Analyse* auswirken, nicht jedoch auf die Bewertung.

CT-Dispositionen wie *Aufgeschlossenheit, Analysefähigkeit, Systematik* und *Selbstvertrauen* wurden ebenfalls in allen Modulen angesprochen. Dennoch zeigten die Erhebungen vor und nach dem Test keine signifikanten Veränderungen in den CT-Dispositionen der Studierenden. Darüber hinaus wurde ein Rückgang der Durchschnittswerte der meisten Subskalen festgestellt. Dies war auf den ersten Blick überraschend, da einige Studien unter Universitätsstudenten signifikante Verbesserungen der CT-Dispositionen gezeigt haben (z. B. [50]). Die in der letzteren Studie berichteten positiven Effekte könnten auf die Besonderheit des Themas zurückzuführen sein oder mit Unterschieden im Instrument zur Beurteilung der CT-Dispositionen zusammenhängen. Überraschender ist, dass auch ein geringer Rückgang der Mittelwerte der meisten Subskalen festgestellt wurde, auch wenn dieser nicht signifikant ist. Dies könnte vor allem dann der Fall sein, wenn die Studierenden ein relativ hohes Ausgangsniveau an Dispositionen aufwiesen (> 3,5 auf der 7-Punkte-Likert-Skala). Außerdem erfordert die Veränderung von eingefahrenen Denkgewohnheiten und Denkmustern viel Zeit und wiederholte Übung, um sie erfolgreich zu ändern.

Diese Ergebnisse waren zwar nicht die erwarteten, aber auch nicht überraschend. Die Studierenden erzielten relativ hohe Werte im Vortest (die Mittelwerte der Subskalen lagen deutlich über 3,5 auf einer 7-Punkte-Likert-Skala), was darauf hindeutet, dass die sie vor den Interventionen ein hohes Maß an CT-Dispositionen hatten. Außerdem erfordert die Veränderung etablierter Denkgewohnheiten und Denkmuster Zeit und wiederholtes erfolgreiches Üben (siehe [51]). Dementsprechend stellt Halpern [51] fest: "*Es scheint klar zu sein, dass die Fähigkeit, klar zu denken, und die Bereitschaft, sich auf den anstrengenden Prozess des Denkens einzulassen, die kritischsten Komponenten der Ausbildung [der Studierenden] sind. Die Förderung der Fähigkeiten zum kritischen Denken ist auch die anspruchsvollste und persönlich lohnendste Aufgabe, mit der sich Psychologen und Pädagogen beschäftigen können*" (S.455).

Außerdem erfordert die Änderung eingefahrener Denkgewohnheiten und Denkmuster viel Zeit und wiederholte Übung, um sie erfolgreich zu ändern. Darüber hinaus sind die

Dispositionen von einem Motivationsfaktor abhängig, der bestimmt, ob Einstellungen gezeigt werden [48]. Die erzielten Ergebnisse stehen im Einklang mit einigen Studien, die über geringe und sogar nicht signifikante Veränderungen der CT-Dispositionen berichteten, was mit Unterschieden in der Sensitivität der Erhebungen zur Erkennung kleiner Veränderungen in den Einstellungen in Verbindung gebracht wurde, oder mit der Tatsache, dass die Interventionen zwar zu kleinen Steigerungen der CT-Dispositionen geführt haben könnten, diese aber nicht groß genug waren, um statistisch signifikant zu sein [48]. Im Gegensatz dazu zeigten andere Studien eine deutlich positive Auswirkung von Interventionen, die auf die Entwicklung von CT bei Hochschulstudenten abzielten [52]. Man muss sich jedoch darüber im Klaren sein, dass es schwierig ist, Vergleiche zwischen Studien anzustellen, die über die Nützlichkeit von pädagogischen Interventionen zur Förderung von CT-Dispositionen bei Studierenden berichten, da die zur Bewertung verwendeten Instrumente in den einzelnen Studien oft sehr unterschiedlich sind und unterschiedliche Gruppen von CT-Dispositionen betreffen. In der vorliegenden Studie scheint ein Semester CTBAC nicht auszureichen, um eine deutliche, signifikante positive Veränderung der CT-Dispositionen zu erreichen.

## Griechenland

Die University of Western Macedonia (UOWM) hat in Zusammenarbeit mit der Elementary Experimental School of Florina (EESF) im akademischen Jahr 2021-2022 drei Critical Thinking Apprenticeships Curricula (CTAC) im Fachbereich Lehrerbildung durchgeführt. Insbesondere wurden im Wintersemester 2021-2022 drei Module durchgeführt, und die Studierenden, die diese Module besuchten, bildeten die Experimentalgruppe für die in IO4 beschriebene Intervention. Bei den drei Modulen handelte es sich um "Didaktik der Biologie" (n=83), "Didaktik der Naturwissenschaft" (n=62) und "Didaktik der Umweltkunde" (n=12). Bei allen drei Modulen handelt es sich um Wahlpflichtkurse, die von der Abteilung für Grundschulpädagogik der UOWM angeboten werden, mit Ausnahme des Moduls "Didaktik der Biologie", der den Studierenden von der Abteilung für frühkindliche Bildung der UOWM angeboten wird. Darüber hinaus wurde im Frühjahrssemester 2021-2022 das Modul "Didaktik

der Naturwissenschaft" mit neuen Studierenden (n=85) durchgeführt, die die Kontrollgruppe der aktuellen Studie bildeten. Alle Module wurden über einen Zeitraum von 13 Wochen durchgeführt. Die Module wurden einmal pro Woche mit einer Dauer von drei Stunden durchgeführt. Für jedes Modul wurde also eine dreistündige Intervention pro Woche durchgeführt. Wie in der Konzeption des CTAC von Mäkiö und Kollegen [1] für die beiden Module "Didaktik der Naturwissenschaft" und "Didaktik der Umweltkunde " beschrieben, war der erste Teil des Moduls theoretisch (ca. 5 Wochen), und dann entwarfen die Lehramtsstudenten ihre Lehr-Lernsequenzen und setzten sie in den Grundschulen um. Im Gegensatz dazu entwarfen die Lehramtsstudenten, die das Modul "Didaktik der Biologie" besuchten, ihre Lernsequenzen für den Unterricht; diese wurden jedoch nie im schulischen Kontext umgesetzt.

Die drei Module wurden wie in Mäkiö und Kollegen [1] beschrieben umgesetzt. Im Modul "Didaktik der Umweltkunde " wurden die Fallstudien jedoch in der Klasse durchgeführt und nicht, wie ursprünglich vorgeschlagen, in Moodle. Diese Abweichung vom ursprünglichen CTAC-Konzept ist auf die geringe Anzahl von Studierenden zurückzuführen, die sich für die Kurse entschieden haben, so dass der Dozent mehr Zeit für Diskussionen und Anwendungen im Unterricht aufwenden konnte. Während der Durchführung des CTAC erwies es sich für die Ausbilder als recht schwierig, die Gesamtzahl der Lehrerstudenten in die Vortest-Messungen der CT-Fähigkeiten und -Dispositionen einzubeziehen. Von den insgesamt 157 Studierenden, die alle drei Module besuchten, haben nur 63 beide Messungen abgeschlossen. Die drei Ausbilder berichteten über keine weiteren Probleme bei der Durchführung der Module.

Sowohl die UOWM als auch die EESF erkannten die Bedeutung bestimmter Fähigkeiten und Dispositionen, die die Lehramtsstudenten während des Praktikums verbessern sollten. Daher konzentrierten sich die Critical Thinking Blended Apprenticeships Curricula mehr auf spezifische Fähigkeiten wie *Analyse*, *Schlussfolgerung*, *Bewertung* und *Selbstregulierung*. Die Ergebnisse der Vor-/Nachtests deuten darauf hin, dass sich die Intervention insgesamt positiv auf alle Fähigkeiten auswirkt, auch auf solche, die bei der Ausarbeitung der Lehrpläne und Aktivitäten nicht ausdrücklich angesprochen wurden (z. B. *Interpretation* und *Erläuterung*).

Obwohl nicht explizit angestrebt, war die Fähigkeit, die wesentlichen und unwesentlichen Faktoren einer bestimmten Situation zu erkennen (d. h. *Interpretation*) und die Argumentation zu begründen (d. h. *Erläuterung*), ebenfalls Teil der Aktivitäten, mit denen Lehrer und Studierende während des Semesters beschäftigt waren. Darüber hinaus berichteten die Kursleiter nach der Einführung des Critical Thinking Blended Apprenticeships Curricula von einer relativ guten Beteiligung und einer wahrgenommenen Verbesserung der CT-Fähigkeiten und CT-Dispositionen der Studierenden. Obwohl die Dozenten Veränderungen in den CT-Fähigkeiten während des Semesters wahrnahmen, erwarteten einige von ihnen statistisch signifikante Veränderungen in einigen Fähigkeiten (z. B. *Reflexion* für den Kurs "Didaktik der Naturwissenschaft") aufgrund des Fokus der Dozenten auf oder der wahrgenommenen Wichtigkeit von spezifischen Fähigkeiten und Dispositionen. An anderer Stelle wird beschrieben, dass der Kurs "Didaktik der Naturwissenschaft" die Dimensionen des reflektierenden Denkens bei den Lehramtsstudenten erhöhte [53]. Es wurde jedoch ein anderer Ansatz zur Messung der Reflexion verwendet, um den entsprechenden Unterschied zu ermitteln. Insgesamt lassen unsere Ergebnisse den Schluss zu, dass die Lehramtsstudenten während des Semesters ihre CT-Fähigkeiten verbessert haben.

Was die Dispositionen anbelangt, so wurden die Aktivitäten so konzipiert, dass sie *Aufgeschlossenheit*, *Systematik*, *Selbstvertrauen*, *Neugier* und *kognitive Reife* fördern. Andererseits ergaben die Vor-/Nachtests keine statistisch signifikante Veränderung in den CT-Dispositionen der Lehrerstudenten. Darüber hinaus stellten wir einen Rückgang des Mittelwerts der Subskala Aufgeschlossenheit fest. Diese Ergebnisse waren zwar nicht die erwarteten, aber auch nicht überraschend. Erstens erzielten die Lehramtsstudenten bei der Vormessung relativ hohe Werte (der Mittelwert der Subskalen lag über 3,5 auf einer 7-Punkte-Likert-Skala), was darauf hindeutet, dass die oben genannten Dispositionen bereits vorhanden sind und verbessert werden könnten. Allerdings werden die CT-Dispositionen als relativ stabil angesehen und bedürfen daher möglicherweise einer systematischen und kontinuierlichen Anstrengung zu ihrer weiteren Verbesserung, die zahlreiche Kurse während des vierjährigen Grundstudiums umfasst. Darüber hinaus stimmen diese Ergebnisse mit ähnlichen Studien in

der Literatur überein, in denen nur wenige statistisch signifikante Veränderungen in den CT-Dispositionen der Studenten festgestellt wurden, einschließlich eines Rückgangs der Dispositionen (siehe [54, 55]). In Anbetracht der Tatsache, dass Dispositionen die Bereitschaft und Tendenz zu kritischem Denken anzeigen, können sie als noch wichtiger als Fähigkeiten angesehen werden. Die Kultivierung von Dispositionen scheint jedoch eine anspruchsvolle Aufgabe zu sein, die durch eine Kultur des kritischen Denkens unterstützt werden kann, die einen Unterricht in dieser Richtung beinhaltet. Im Einzelnen werden in einem solchen Unterricht Aktivitäten hervorgehoben, die sich sowohl auf die Emotionen als auch auf die Kognition konzentrieren, und zwar im Hinblick auf die Interaktion mit sorgfältigen und kontinuierlichen Übungen zum CT [56]. Diese Art von Ansatz ist wahrscheinlich zeitaufwändiger und erfordert mehr Zeit als ein Semester und sollte im gesamten Lehrplan verankert werden.

Im Allgemeinen werden die Unterschiede in den CT-Fähigkeiten und -Dispositionen, die zwischen den beiden untersuchten Modulen "Didaktik der Naturwissenschaft" und "Didaktik der Biologie" festgestellt wurden, auf die unterschiedliche Gestaltung der Module zurückgeführt. Zu den auffälligsten Unterschieden gehörten die unterschiedlichen didaktischen Ansätze der Dozenten sowie die Umsetzung der TLS der Studierenden in realen Kontexten für das Modul "Didaktik der Naturwissenschaft" (d. h. in der Schule).

Außerdem wurden die Auswirkungen der Maßnahmen zwischen der Versuchs- und der Kontrollgruppe verglichen. Die Ergebnisse zeigten, dass die Versuchsgruppe einen größeren Zuwachs an integrierten Ergebnissen sowohl bei den Fähigkeiten als auch bei den Dispositionen hatte. Dieses Ergebnis ist plausibel, wenn man bedenkt, dass in allen drei Kursen der Versuchsgruppe zu Beginn des Semesters expliziter CT-Unterricht durchgeführt wurde. Darüber hinaus haben frühere Forschungen gezeigt, dass explizite CT-Instruktion die Studierenden begünstigt (z. B. [57]). Darüber hinaus wiesen Heijltjes und Kollegen [58] darauf hin, dass die Kombination von explizitem Unterricht und Üben für die Studierenden vorteilhafter ist als andere Unterrichtsansätze, wie z. B. die implizite Unterweisung von CT. Bei der Untersuchung der einzelnen Fähigkeiten und Fertigkeiten zeigten unsere Ergebnisse

jedoch, dass sich nur einige Fähigkeiten zugunsten der Versuchsgruppe verbesserten, z. B. "Bewertung", "Schlussfolgerung" und "Erläuterung". Diese Fähigkeiten gehörten zu den erwarteten CT-Ergebnissen des Kurses "Didaktik der Naturwissenschaft". Obwohl sich der Kurs auch auf die CT-Fähigkeiten "Analyse" und "Selbstregulierung" konzentrierte, wurde bei diesen Fähigkeiten keine Verbesserung festgestellt. Wir können argumentieren, dass spezifische CT-Fähigkeiten zugunsten der Versuchsgruppe verbessert wurden, weil der Ausbilder während des Unterrichts mehr auf diese Fähigkeiten durch die Aktivitäten, an denen die Studierenden beteiligt waren, beharrte. Zur Veranschaulichung: Nach der Umsetzung ihrer Lehr-Lern-Sequenzen (TLS) präsentierten die Lehramtsstudenten ihre ursprünglichen Entwürfe, die TLS, die sie tatsächlich umgesetzt hatten, und reflektierten über die CT-Fähigkeiten und CT-Dispositionen, die sie einsetzten, sowie über die Änderungen, die sie vornehmen würden, um ihre TLS erfolgreicher zu gestalten. Was die einzelnen Dispositionen betrifft, so zeigten unsere Ergebnisse, dass sich in der Versuchsgruppe nur die Aufmerksamkeit verbessert hat, d. h. die Bereitschaft der Studierenden, sich zu fokussieren und zu konzentrieren, sich der Umgebung, des Kontexts, der Konsequenzen und möglicher Hindernisse bewusst zu sein. Dieser Befund lässt sich damit begründen, dass die Studierenden der Versuchsgruppe vor der Durchführung ihrer TLS aufgefordert wurden, ein Video zu erstellen, das zeigt, dass die Experimente, die sie im Unterricht durchführen würden, realistisch sind und die erwarteten Ergebnisse im Hinblick auf den Wissenserwerb der Studierenden liefern. Damit die Lehrkräfte bei dieser Aktivität erfolgreich sein konnten, mussten sie den Kontext, das schulische Umfeld sowie die Folgen berücksichtigen, die das Experiment für die Studierenden haben könnte, je nachdem, wie erfolgreich es durchgeführt werden würde. Diese Aktivität wurde im Kurs der Kontrollgruppe nicht durchgeführt. Daher gehen wir davon aus, dass der Unterschied in dieser Disposition zu erwarten war. Insgesamt kann bei der Betrachtung des durchgeführten Kurses für die Experimental- und die Kontrollgruppe davon ausgegangen werden, dass die unterschiedlichen durchgeführten Aktivitäten sowie die explizite Vermittlung von CT der Grund für die festgestellten Veränderungen zwischen den beiden Gruppen sowohl in Bezug auf die CT-Fähigkeiten als auch in Bezug auf die CT-Dispositionen sein könnten.

Aufgrund der Anzahl und der Länge der geplanten Aktivitäten (sie waren Teil des Wochenplans der Einheit) erwarteten wir eine stärkere Entwicklung der CT-Fähigkeiten und -Dispositionen zu beobachten. Die Planung und Durchführung der Aktivitäten war jedoch mit einigen Einschränkungen verbunden. Erstens war das Wintersemester 2021-2022 das erste Semester, in dem in Griechenland seit der Ausrufung der Pandemie Präsenzunterricht stattfand, so dass wir Bedenken hinsichtlich der Bereitschaft der Lehrenden und Lernenden hatten, sich systematisch an den Unterricht zu halten und an allen geplanten Aktivitäten teilzunehmen. Die Kursleiter berichteten jedoch von einer relativ guten Beteiligung der Lehrkräfte am Unterricht, trotz der Schwierigkeiten (z. B. beim Ausfüllen der Vor-Nach-Messungen, beim Einhalten einiger Fristen, beim Erholen von Covid-19 usw.). Eine weitere Einschränkung liegt in der Art der Aktivitäten. Die meisten Aktivitäten fanden in den Klassen während des Unterrichts statt und erforderten die aktive Beteiligung von Lehrern und Studierenden (z. B. kritische Diskussion, Brainstorming, Arbeit in Gruppen usw.). Diese Aktivitäten könnten einige - introvertierte - Studierende, die Schwierigkeiten mit dem Sprechen in der Öffentlichkeit haben oder es vermeiden, ihre Meinung in der Klasse zu äußern, trotz des guten und unterstützenden Klimas, über das die Lehrkräfte berichteten, nicht aktiv einbeziehen. Vor diesem Hintergrund sind die Studierenden möglicherweise auch nicht mit dieser Art von interaktivem Unterricht vertraut, bei dem ihre aktive Teilnahme in gewisser Weise obligatorisch ist und im Unterricht ständig gefördert wird. Daher benötigen einige von ihnen möglicherweise mehr Zeit, um sich an den Prozess des aktiven Lernens anzupassen und sich aktiv daran zu beteiligen.

## Litauen

Die Universität Vilnius führte einen CTBAC ein - *Englisch für akademische Zwecke und Forschung* - ein obligatorischer Englischkurs für bestimmte Zwecke, der vom Institut für Fremdsprachen der Philologischen Fakultät der Universität Vilnius angeboten wird. Die Umsetzung des CTBAC erfolgte in Zusammenarbeit mit dem Sprachenzentrum für den öffentlichen Dienst im Rahmen des Studiengangs *Internationale Beziehungen und*

*Politikwissenschaft* (BA) an der Universität Vilnius; er wurde den Studierenden des ersten Studienjahres des Studiengangs *Internationale Beziehungen und Politikwissenschaft* (BA) am Institut für Internationale Beziehungen und Politikwissenschaften der Universität Vilnius angeboten. Der Lehrplan wurde im akademischen Jahr 2021-2022 (zwischen September und Mai) eingeführt, wie von Mäkiö und Kollegen beschrieben [1]. Ursprünglich war ein weiteres CTBAC unter dem Namen *Englisch für akademische Zwecke und Forschung* geplant, das im Fach *Kindheitspädagogik* eingeführt werden sollte. Es wurde im Herbstsemester des akademischen Jahres 2020/2021 erfolgreich mit den Studierenden des ersten Studienjahres der Philosophischen Fakultät der Universität Vilnius durchgeführt. Die Studierenden wurden gleich zu Beginn von CTBAC einem Vortest unterzogen, aber leider nahmen sie nicht am Nachtest teil. Insbesondere wurde den Studierenden eine editierbare Version des Fragebogens zur Verfügung gestellt, die es ihnen ermöglichte, den Inhalt und die Skalen der Items zu ändern; aus diesem Grund wurden ihre Antworten im Nachtest als nicht gültig betrachtet, sie wurden nicht statistisch ausgewertet und von der abschließenden Bewertung und dem Abschlussbericht ausgeschlossen.

Die Gesamtdoktrin des Lehrplans basiert auf aufgaben- und handlungsorientierten Ansätzen (AoA) in Anlehnung an die aktualisierte Version des Gemeinsamen Europäischen Referenzrahmens für Sprachen: Lernen, Lehren, Beurteilen [59] und einer neuen Vision des Fremdsprachenunterrichts, wie sie von Piccardo und North [60] beschrieben wurde. Dieser Ansatz wurde vom Institut für Fremdsprachen der Philologischen Fakultät der Universität Vilnius im Fremdsprachenunterricht ab dem Studienjahr 2019/2020 systematisch umgesetzt [61]. Der handlungsorientierte Ansatz stützt sich auf mehrere Säulen, die den Kern des CTAC bilden: der Lernende als sozialer Akteur, sprachliche Aktivitäten in einem bestimmten sozialen Kontext und reale, handlungsorientierte Aufgaben [60]. Folglich war der CTBAC-Lehrplan schülerzentriert, und die für die Studierenden konzipierten Aufgaben waren handlungsorientiert, sollten authentisch sein und Problemlösungen sowie Situationen aus dem wirklichen Leben beinhalten.

Im Rahmen des Lehrplans sollten sich die Studierenden mit den Anforderungen des wissenschaftlichen Schreibens vertraut machen. Zu diesem Zweck wurden zwei verschiedene Aufgaben vorgeschlagen. Erstens erwarben sie theoretische Kenntnisse über die gültige Struktur wissenschaftlicher Forschung. Nachdem sie das Genre hinreichend verstanden hatten, sollten sie einen Forschungsvorschlag einreichen, um die nächste Aufgabe vorzubereiten und zu antizipieren. Die Schreibaufgabe zielte auch auf die Entwicklung der CT-Fähigkeiten der Studierenden ab. Die Studierenden mussten Forschungsartikel analysieren und darauf aufbauend Einblicke in mögliche zukünftige Forschungen geben und einen Plan für ihre eigene Forschung vorschlagen. Die Aufgabe zielte also auf die CT-Fähigkeiten *Analyse, Erläuterung, Interpretation, Schlussfolgerung* und *Bewertung* ab. Zweitens wurde am Ende des ersten Semesters eine internationale Konferenz simuliert, zu der jeder Student mit einer individuellen Präsentation beitragen sollte. Die gleiche Aktivität wurde am Ende des 2. Semesters organisiert, nur dass die Studenten diesmal in Teams arbeiteten. Die Aktivität war als offene Veranstaltung strukturiert, an der andere Studierende oder Lehrkräfte teilnehmen konnten. An beide Präsentationen schloss sich eine Fragerunde an, in der die Vortragenden auf die Fragen, Kommentare und Vorschläge des Publikums eingingen. Die Aufgaben gingen von einem klar definierten, authentischen Szenario aus, das auf den in der Kursbeschreibung genannten allgemeinen Themen basierte. In Bezug auf den CT wurden in den Studien die oben genannten intellektuellen Eigenschaften getestet. Die Einzel- und Teampräsentationen, die die Studierenden durchführen mussten, zielten auf die Verbesserung der CT-Fähigkeiten und -Dispositionen ab. Die Studierenden mussten die aus den von ihnen zu analysierenden Forschungsartikeln gewonnenen Daten objektiv und zusammenhängend präsentieren. Wir erwarteten von den Studierenden eine kritische Analyse der bisherigen Forschungsergebnisse, die Identifizierung und Definition von Schlüsselkonzepten innerhalb eines ausgewählten theoretischen Rahmens, den Vergleich von Forschungsergebnissen aus mindestens einigen Forschungsstudien, die Angabe konkreter Beispiele, das Ziehen von Schlussfolgerungen und das Aufzeigen verschiedener Implikationen. Von großer Bedeutung war auch die Fähigkeit der Studierenden, die Forschungsergebnisse aus den Artikeln zu vergleichen und sie mit dem sozialen, politischen oder wirtschaftlichen Kontext in Beziehung

zu setzen. Wir können also feststellen, dass die Entwicklung der CT-Fähigkeiten der Studierenden (*Analyse, Interpretation, Schlussfolgerung, Vergleich*) während des CTBAC ausdrücklich im Mittelpunkt stand.

Die Analyse der Fortschritte, die mit der Einführung der CTBACs verbunden sind, beweist die Entwicklung der CT der Studierenden. Die durchgeführten Maßnahmen führten zu einem Anstieg der Punktzahl für die integrierten CT-Fähigkeiten um mehr als 10 Punkte sowie zu einem signifikanten Fortschritt in allen CT-Teilfertigkeiten mit Ausnahme der Bewertung. Bei dieser Teilfertigkeit war die Punktzahl nach dem Test zwar höher als vor dem Test, aber der Unterschied war nicht signifikant.

Dennoch zeigten die Vergleiche zwischen Vor- und Nachtests, dass die CT-Dispositionen durch die Pilotmaßnahmen nicht signifikant verbessert wurden. Dort, wo ein Fortschritt zu beobachten war, ist er noch nicht signifikant. Dies kann verschiedene Gründe haben. Erstens die Tatsache, dass es schwieriger ist, Einstellungen zu ändern als Verfahren. Einstellungen erfordern ein tieferes Verständnis und Engagement für die Grundsätze und Werte des CT und können durch die Stimmung, Interessen und Entscheidungen des Einzelnen stärker beeinflusst werden [62]. Teilweise wird dies auch durch subjektive, unvorhersehbare Bedingungen erklärt. Der Nachtest wurde gegen Ende des akademischen Jahres durchgeführt, wenn normalerweise Stress und die Fähigkeit, sich auf bestimmte Aufgaben zu konzentrieren, nachlassen. Um das Risiko zu mindern, könnten wir eine alternative Strategie in Betracht ziehen. Der Wettbewerb nimmt im 2. Semester zu, wenn die Studierenden Noten anstelle von "bestanden/nicht bestanden"-Aufgaben (1. Semester) erhalten. Traditionell stellen Noten einen Anreiz dar, könnten aber zusätzlichen Druck auf die Studierenden ausüben, mit sichtbaren Folgen für ihre Bereitschaft, ihre CT-Dispositionen durchzusetzen und zu nutzen. Die Hypothese wurde nicht getestet, und wir haben nicht genügend Daten, um sie zu stützen. Die erhobenen Daten deuten darauf hin, dass sich die CT-Dispositionen gegen Ende des Studienjahres leicht verschlechterten. Die Ergebnisse scheinen darauf hinzudeuten, dass der getestete Lehrplan eine bessere Umverteilung der Aufgaben über die beiden Semester hinweg erfordert, um das Interesse der Lernenden an den untersuchten Themen und Aufgaben

aufrechtzuerhalten. Insgesamt lassen unsere Ergebnisse den Schluss zu, dass die Studierenden während des Semesters ihre CT-Fähigkeiten verbessert haben.

## Portugal

Die Pilotmodule wurden wie in IO3 [1] vorgeschlagen durchgeführt, wenn auch mit kleinen Unterschieden. Im Bereich der *Bildgebung* wurden nur zwei der drei vorgeschlagenen Maßnahmen mit den Studenten durchgeführt, da sich die Anzahl der nationalen Feiertage mit dem Unterricht überschneidet (insgesamt fielen zwei Wochen Unterricht aus). In den Modulen *Deontologie* und *Gynäkologie, Andrologie* und *Geburtshilfe* wurden alle drei geplanten Aktivitäten durchgeführt. Im ersten Modul bestanden die Aktivitäten aus der Analyse einer dilemmatischen Situation, gefolgt von der Äußerung und Diskussion verschiedener Standpunkte [1], während sich die Aktivitäten im zweiten Modul auf die Analyse eines klinischen Zustands, das Screening von Differentialdiagnosen und die Entscheidungsfindung über die beste Vorgehensweise bei einem bestimmten Zustand bei einem Tier mit besonderen Merkmalen konzentrierten [63].

In allen drei Pilotkursen zogen es alle Studierenden vor, die Pilot-CTBACs als Lernmethode zu verwenden, anstatt eine Kontrollgruppe zu bilden. Die Bildung einer Kontrollgruppe außerhalb der Module war nicht möglich, da diese Kurse in einem Studienjahr nur einmal angeboten werden. Der Inhalt und der Hintergrund der veterinärmedizinischen Studiengänge an anderen Universitäten waren unterschiedlich, und es wurde beschlossen, keine Kontrollgruppe mit externen Studierenden zu bilden. Daher wurde beschlossen, eine kleine Kontrollgruppe zu bilden, die sich aus Studenten zusammensetzte, die am *Hospital Veterinário do Atlântico* in der Ausbildung stehen.

Es war schwierig, die meisten Studenten dazu zu bewegen, die drei Fragebögen zu beantworten, wie es während der Durchführung der CTBACs geplant und verlangt worden war, und im Laufe der Zeit haben sich die Studenten den CT-Fragebögen entzogen. Aus diesem

Grund liegt die Repräsentativität der Studierenden bei den Vor- und Nachtests oft unter 50 % der Studierenden, die an den Aktivitäten teilgenommen haben.

Die Ergebnisse der durchschnittlichen Zuwächse zeigten, dass die Pilotinterventionen in den CTBACs zu Veränderungen in den CT-Fähigkeiten und CT-Dispositionen beitrugen. Im Allgemeinen wirkte sich das Geschlecht nicht auf die Ergebnisse der CT-Fähigkeiten und CT-Dispositionen aus, mit Ausnahme von *Analyse* und *Organisation*, wobei erstere bei Männern höhere Durchschnittswerte aufwies als bei Frauen, und letztere ein umgekehrtes Muster zeigten. Zuwächse bei den CT-Fähigkeiten wurden in allen drei Pilotkursen verzeichnet, ebenso wie bei drei der sechs CT-Dispositionen (*Reflexion*, *Durchhaltevermögen* und *intrinsische Motivation*) und der integrierten Bewertung der Dispositionen.

Dem allgemeinen Muster folgend, waren die durchschnittlichen Zuwächse bei den integrierten CT-Fähigkeiten höher (fast 7,2 Punkte) als bei den Dispositionen (etwa 1 Punkt). Die Zuwächse bei den CT-Fähigkeiten nach den CTBACs stimmen eng mit den Fähigkeiten überein, die als Ergebnisse für die in IO3 [1, 64] beschriebenen Lernstrategien vorgeschlagen wurden, was auf eine korrekte Ausrichtung der Interventionen auf die geplanten CT-Ergebnisse hindeutet. Ein ähnliches Ergebnis wurde nicht für die CT-Dispositionen gefunden, was Anlass zum Nachdenken gibt.

Die Daten der portugiesischen Studenten zeigten, dass die für den Studiengang Veterinärmedizin konzipierten Interventionen zu einem Anstieg einiger der vom Arbeitsmarkt gelobten CT-Dispositionen führten, nämlich *Reflexion*, *Durchhaltevermögen* und *intrinsische Motivation*. Während das Konstrukt der erstgenannten Disposition in der SENCTDS-Skala eng mit der Facione-Konzeptualisierung von CT-Dispositionen und dem Verständnis einer reflektierenden, skeptischen Haltung übereinstimmt, umfassen die beiden letztgenannten Dispositionen eine Mischung aus verschiedenen Haltungen, die in neuen Konzepten kombiniert werden. Nach der Konzeptualisierung von Quinn et al. [3] stellen diese Dispositionen positive Eigenschaften oder Einstellungen dar, die sowohl in akademischen Kontexten als auch in der Arbeitswelt erforderlich sind. *Durchhaltevermögen* steht für

Widerstandsfähigkeit, die Motivation, bei anspruchsvollen Aufgaben durchzuhalten, gute Leistungen im Job zu erbringen und den Wunsch, voranzukommen. Die *intrinsic Motivation* steht für die Fähigkeit, einer Aufgabe oder einem Problem oder dem Lernprozess und der Suche nach Lösungen positiv und enthusiastisch gegenüberzustehen; sie umfasst auch den inneren Antrieb, unabhängig von jeglicher Belohnung nach Antworten zu suchen [3]. In den vorgeschlagenen Interventionen [1] wurden die gezielten CT-Dispositionen entsprechend der Facione-Konzeptualisierung der CT-Fähigkeiten und -Dispositionen identifiziert. Allerdings wurden nur einige von ihnen mit dem im Projekt verwendeten Instrument als solche bewertet (z. B. *Reflexion*, *Achtsamkeit* und *Aufgeschlossenheit*). Von diesen wurde nur die *Reflexion* innerhalb der CTBACs positiv stimuliert.

*Durchhaltevermögen*, verstanden als die Neigung, über das eigene Verhalten oder die eigene Motivation nachzudenken, wird mit einer besseren Entscheidungsfindung in der Praxis und der Fähigkeit, fundierte Urteile zu fällen, in Verbindung gebracht [3]. Andererseits spiegelt die *intrinsic Motivation* die Neugier, die Beherrschung und die intrinsische Zufriedenheit des Lernprozesses wider [3] und damit die selbstregulierende Einstellung, die wir bei Hochschulstudenten entwickeln möchten [65].

Obwohl es bei den Studierenden vor der Durchführung der Pilotmodule Unterschiede in bestimmten CT-Fähigkeiten gab (nämlich in den Bereichen *Analyse* und *Erläuterung*), die bei den Studierenden, die in den Modulen in den späteren Jahren des Studiengangs Veterinärmedizin eingeschrieben waren, höhere Werte aufwiesen, gab es beim Vergleich der Zuwächse in den CT-Fähigkeiten und den entsprechenden Dimensionen keine Unterschiede zwischen den drei Modulen. Es gab auch keine Unterschiede bei den durchschnittlichen Zuwächsen der Studierenden in den drei Module. Dieses Ergebnis war überraschend, denn obwohl ein allgemeiner Rahmen verwendet wurde, unterschieden sich die Komplexität der Aktivitäten, die Strategien, die Form ihrer Umsetzung und die für die Aktivitäten vorgeschlagenen Ergebnisse in den drei CTBACs.

Diese Ergebnisse deuten darauf hin, dass die Intention der Intervention mehr als die Art der Strategie die CT der Studierenden effizient verbessern kann. Das Hauptproblem könnte darin bestehen, den Schwerpunkt vom Produkt (Wissen) auf den Prozess (Argumentation) zu verlagern und damit zu einem konstruktivistischen Lernansatz überzugehen.

## Rumänien

Während der drei in Rumänien durchgeführten Kurse (*Pädagogik und Didaktik der Finanzbuchhaltung, Virtuelle Lernumgebungen in der Wirtschaft und Unternehmenskommunikation*) präsentierten Ausbilder von Arbeitsmarktorganisationen den Studenten verschiedene Fallstudien aus der Praxis. Auf diese Weise konnten die Studierenden auf der Grundlage der zuvor erworbenen theoretischen Informationen auch die Endergebnisse in der Praxis sehen, nachdem sie die zugewiesenen Szenarien analysiert und interpretiert hatten.

In einem Lernszenario im Modul *Pädagogik und Didaktik der Finanzbuchhaltung* mussten die Studierenden sowohl Lernende als auch Lehrende sein und die zu vermittelnden Inhalte identifizieren und analysieren. Sie mussten auch interaktive Materialien/Präsentationen erstellen und recherchieren, welche Lehrmethoden je nach dem zugewiesenen Thema am besten geeignet waren, um die Aufmerksamkeit ihrer Kollegen zu gewinnen [1].

Im Modul "*Virtuelle Lernumgebungen in der Wirtschaft*" lernten die Studierenden, wie sie interaktive Plattformen erstellen können, die es ihnen ermöglichen, Bildungsaktivitäten auf hohem Niveau durchzuführen [1]. Außerdem wurden sie von Ausbildern der Arbeitsmarktorganisationen in verschiedenen Lehrmethoden und Software (z. B. Canva, Google-Sites) unterrichtet, so dass sie in der Lage sein werden, ihren Unterricht in naher Zukunft auf hohem Qualitätsniveau durchzuführen.

Im dritten Modul „*Unternehmenskommunikation*“ lernten die Studierenden verschiedene theoretische Begriffe über den Kommunikationsprozess (Techniken, Kanäle, Grenzen/Einschränkungen usw.) [1]. Sie wurden in verschiedene Situationen versetzt:

Schwierigkeiten bei der Umsetzung eines Projekts, Identifizierung von Hindernissen in einer Geschäftsbesprechung, Analyse von Dokumenten, um die Fähigkeit zum kritischen Denken zu entwickeln.

In Anbetracht der Tatsache, dass die meisten Lernszenarien, an denen die Studenten beteiligt waren, sie mit der Analyse und Interpretation von Informationen beschäftigten, gab es signifikante Veränderungen bei den *Interpretationsfähigkeiten* und dem *Durchhaltevermögen* der Studenten aufgrund der Interventionen der Ausbilder der Arbeitsmarktorganisationen. Darüber hinaus ist es den Studenten in den drei durchgeführten Kursen gelungen, bestimmte Dispositionen (auf einem höheren oder niedrigeren Niveau) zu entwickeln, wie z.B. Aufmerksamkeit, Aufgeschlossenheit, intrinsische Zielmotivation, und zwar aufgrund von Fallstudien, die in der Praxis anwendbar waren. Auf diese Weise konnten die Studierenden eine globale Perspektive einnehmen: von theoretischen Begriffen bis hin zu den Ergebnissen im wirklichen Leben. In Anbetracht der Tatsache, dass die Intervention der Arbeitsmarktausbilder insgesamt einen positiven Einfluss auf die Entwicklung der Fähigkeit der Studierenden zu kritischem Denken hatte, wird empfohlen, die von den Lehrkräften in den Hochschuleinrichtungen angewandten Lehrmethoden/Techniken mit den von den Arbeitsmarktvertretern geförderten zu harmonisieren

### TEIL III – THINK4JOBS-LEITLINIEN FÜR DIE UMSETZUNG VON CTBACS

Die Umsetzung der CTBACs erfolgte im ersten akademischen Jahr nach den Anpassungen der Covid-Pandemie, im Herbst- und Frühjahrssemester 2021-2022. In den meisten Ländern war dies das erste Jahr, in dem Präsenzunterricht seit der Ausrufung der Pandemie stattfand; die Sorgen über die Abriegelung der Hochschulen waren noch in aller Munde. Nichtsdestotrotz zeigte das Engagement der Studierenden, dass sie den Unterricht wieder aufnehmen und an den geplanten Aktivitäten teilnehmen wollten, auch wenn sie mit einer höheren Arbeitsbelastung im Vergleich zum traditionellen Lernprozess rechnen. In einigen Kontexten war die Entwicklung von Aktivitäten in Gruppen ein positiver Faktor, wenn die Aufgabe zu anspruchsvoll erschien.

Insgesamt deuten die Ergebnisse der CTBAC-Pilotprojekte darauf hin, dass sich die Interventionen unabhängig von der Disziplin, in der kritisches Denken angesprochen wurde, als vorteilhaft erwiesen haben, was die Übertragbarkeit der CT-Fähigkeiten und CT-Dispositionen von den Lehrplänen auf die Ausbildung und vom Arbeitsmarkt auf die akademische Welt (von der Ausbildung auf die Lehrpläne) betrifft. Die enge Zusammenarbeit mit LMO verknüpfte den Arbeitsmarkt und das berufliche Umfeld miteinander und brachte einen neuen dynamischen Lehransatz in die Akademie ein, wobei beide Teile vom Einsatz verschiedener aktiver Methoden in den Kursen im Rahmen des Projekts profitierten, wobei ein besonderer Schwerpunkt auf dem entdeckenden Lernen durch eigene Erfahrungen lag [66].

Im Folgenden stellen wir eine Reihe von Leitlinien für die Umsetzung der CT-Lehrpläne für gemischte Ausbildungsgänge vor, die auf den Erfahrungen beruhen, die während der Pilotphase der blended Module gesammelt wurden.

## 1. Erklären Sie, was Sie tun - Warum ist CT auf dem Arbeitsmarkt wichtig?

Es ist wichtig, den Studierenden explizit zu erklären, warum CT eine entscheidende Kompetenz in der heutigen Arbeitswelt ist, die vom Arbeitsmarkt in allen Berufen hoch gelobt wird [8, 67, 68], wobei die Besonderheiten hervorzuheben sind, die sich aus den Unterschieden im Berufsfeld ergeben können [69, 70]. Dies kann erreicht werden, indem man die Erfahrungen des Arbeitsmarktes in den Unterricht einbringt, indem man entweder zu Gesprächen mit gut positionierten Fachleuten einlädt oder den Studierenden die Möglichkeit gibt, die Arbeit der Beteiligten aus erster Hand zu besuchen und zu beurteilen. Der Einsatz von realen Arbeitssituationen zur Schulung von CT-Ansätzen zur Problemlösung ist ein entscheidender Schritt in der Ausbildung zu besseren Fachleuten, die über mehr Erfahrung im Treffen von fundierten Entscheidungen bei der Problemlösung verfügen.

Anschließend ist zu klären, wie die für die CT blended Curricula konzipierte Lehrmethode mit den Tätigkeiten von Fachleuten übereinstimmt, die täglich mit spezifischen Herausforderungen, Situationen oder Problemen zu tun haben, und wie diese Aktivitäten geplant wurden, um die Fähigkeit der Studierenden zu stärken, mit diesen umzugehen und sie selbst zu lösen, um so ihren Einstieg in den Arbeitsmarkt zu erleichtern. Wie von Abrami und Mitarbeitern [71] vertreten, führt die explizite Entwicklung von CT zu höheren Gewinnen, insbesondere wenn ein immersiver Ansatz verwendet wird, wie es bei den implementierten CTBACs der Fall war. Unsere Ergebnisse zeigen, dass inhaltspezifische CT-Fähigkeiten und CT-Dispositionen bei den an CTBACs teilnehmenden Studierenden entwickelt wurden, was durch die frühere Arbeit von Abrami et al [72] unterstützt wurde.

Einige der in den Pilotmokursen verwendeten Strategien bezogen sich auf einen handlungsorientierten Ansatz. Durch die Fokussierung auf den Lernenden als sozialen Akteur, auf die Schülerzentrierung, auf den sozialen oder beruflichen Kontext tauchten die eingesetzten Interventionen die Studierenden in kognitiv herausfordernde Situationen ein

und beschäftigten sie mit Aufgaben aus dem realen Leben, die die Aktivierung allgemeiner Kompetenzen und CT-Fähigkeiten erfordern. Unsere Ergebnisse deuten auch darauf hin, dass die explizite Vermittlung von CT den Erwerb von CT-Fähigkeiten und -Dispositionen durch die Studierenden fördert. In den Critical Thinking Blended Apprenticeships Curricula, die für die aktuellen Interventionen entwickelt wurden, wurde die explizite Unterweisung in CT je nach Programm/Disziplin unterschiedlich umgesetzt, was auf deren Besonderheiten zurückzuführen ist (entweder in den vom Lehrplan des Moduls vorgeschriebenen Qualifikationen oder in der Operationalisierung der Ausbildungsgänge). In den meisten Fällen wurde die Erläuterung des CT-Konzepts und seiner Bedeutung jedoch nur zu Beginn der Kurse mit den Studierenden besprochen. Wir sind der Meinung, dass die Reflexion und Selbstregulierung der Studierenden verbessert werden könnte, wenn die Lehrkräfte dies systematisch in ihre Kurse und die Vermittlung der Inhalte einfließen lassen.

Die in einigen Modulen gesammelten Daten scheinen darauf hinzudeuten, dass es wichtig ist, in einem Kurs mehrfach zu betonen, dass die CT-Fähigkeiten und CT-Dispositionen zusätzliche Ergebnisse des Kurses sind. Die Studierenden könnten von der Lehrveranstaltung die Erwartung haben, dass es sich um eine gewöhnliche Lehrveranstaltung handelt, in der das Hauptaugenmerk auf dem Erwerb von kognitivem Wissen liegt, nicht aber auf der Qualität der Aufgabenerfüllung. Die Studierenden sollten über den Zweck der Aufgaben und die Lernergebnisse informiert werden, die nicht nur die Entwicklung fachbezogener kognitiver Kompetenzen, sondern auch allgemeiner sozialer Kompetenzen umfassen, wobei letztere die Entwicklung von CT-Fähigkeiten und CT-Dispositionen einschließen. Das Bewusstsein der Studierenden für die Bedeutung von CT-Fähigkeiten beim Erlernen eines Faches sollte während des gesamten Kurses geschärft werden. Dies steht im Einklang mit El Soufi und See [73], die gezeigt haben, dass nur die explizite Vermittlung von CT-Fähigkeiten die besten Wirksamkeitsnachweise erbrachte.

Wenn die Entwicklung von CT-Fähigkeiten als entscheidend für die Ausbildung der Studierenden identifiziert wurde, dann sollten spezifische CT-Fähigkeiten und Dispositionen in den für den Kurs oder die Module festgelegten Lernergebnissen identifiziert werden, die

mit dem inhaltlichen Wissen verknüpft sind. Diese sollten dann auch während des Moduls bewertet werden.

## 2. CT muss kontinuierlich und durchgängig unterrichtet werden

Die an diesem Projekt teilnehmenden Studierenden waren eine ausgewählte Gruppe von Studierenden aus fünf verschiedenen Fachrichtungen, die ein spezielles Training erhielten, das sie in die Lage versetzen sollte, ein typisches Problem aus ihrem Berufsalltag zu analysieren und zu entscheiden, welche Maßnahmen erforderlich sind, um es zu lösen. Obwohl die Critical Thinking Apprenticeships Curricula in jedem Programm für die Dauer eines Semesters implementiert wurden, argumentieren wir, dass der Unterricht für CT als kontinuierlicher und durchdringender Prozess organisiert werden sollte, der im gesamten Programm gefördert werden sollte. Er sollte zu Beginn jedes Moduls im Studienplan beginnen und in der abschließenden Lehre gipfeln, die normalerweise am Ende des Grundstudiums stattfindet. In diesem Sinne haben die Studierenden ständig die Möglichkeit, ihre CT-Fähigkeiten und CT-Dispositionen zu trainieren und zu kultivieren.

Dies wäre besonders wichtig für CT-Dispositionen, die eine Verinnerlichung [56] und den Erwerb von Denkgewohnheiten erfordern. Ein explizit auf den CT ausgerichteter Lehrplan sollte nicht nur in spezifischen Modulen (die eher mit der Lehre in Verbindung stehen), sondern in allen Modulen eines Fachs auf Universitätsebene eingeführt und angewendet werden. Auf diese Weise können die Studierenden mit dem Konzept vertraut gemacht werden und es ständig üben. Daher wird von den Studierenden erwartet, dass sie sich beim Eintritt in die Lehre der CT-Fähigkeiten und -Dispositionen bewusst sind, die sie bei der Lösung von Problemen in realen Situationen nutzen können, und dass sie die Möglichkeiten zur Übertragung ihrer erworbenen CT-Fähigkeiten und -Dispositionen auf neue Kontexte maximieren. Am Ende würde die Zeit, die für die Anpassung an die Anforderungen des Arbeitsmarktes benötigt wird, reduziert, der Stress der frischgebackenen Absolventen

verringert, während die Qualität der Arbeit gesichert und das Wohlbefinden der Berufsanfänger erheblich gesteigert würde.

### 3. Zeit für die Entwicklung von CT einplanen

Die CTBACs können als ein Instrument zum Aufbau von Kapazitäten zur Förderung des CT in allen Hochschulabschlüssen betrachtet werden, welches die Integration von CT-Fähigkeiten und CT-Dispositionen bei Hochschulstudenten erleichtern und die von Interessenvertretern und politischen Entscheidungsträgern gemeldeten Kompetenzlücken oder -inkongruenzen abmildern soll.

Damit CTBACs zu Ergebnissen führen, benötigen sie jedoch viel Zeit von Professoren und Studenten, um an den vorgeschlagenen Zielen zu arbeiten. Der Zeitrahmen für die Bildungsinterventionen muss sorgfältig festgelegt werden. Wenn mehr als ein Kurs im selben Jahr des Studiengangs diesen Bildungsansatz verwendet, sollten der Zeitrahmen und die Lernergebnisse gemeinsam betrachtet werden. Wenn die Arbeitsbelastung der Studierenden (und der Lehrkräfte) nicht zu hoch wird, sind sowohl das Engagement als auch der Lernerfolg erwartungsgemäß höher. Zusammen mit der Wiederholung, die durch die Einführung dieses Ansatzes in allen Lehrplänen gefördert wird, würden die kumulativen Gewinne nicht nur für die CT-Fähigkeiten, sondern auch für die CT-Dispositionen durchgängig erzielt werden, da Änderungen in der Einstellung mehr Zeit benötigen, um aufgenommen zu werden.

Für die Entwicklung von CTBACs ist eine sorgfältige Planung erforderlich, da sie sich auf die Entwicklung von übertragbaren Fähigkeiten wie CT konzentrieren. Die Qualität der Ausbildung und die konsequente Verbesserung der Fähigkeiten werden durch die für das Lernen zur Verfügung stehende Zeit beeinflusst. Sie erfordern entweder eine gute Planung oder ein angemessenes Klassenraummanagement und die Fähigkeit, die Zeit für die Aufgaben an die Merkmale der Studierendengruppe anzupassen. Auch wenn der Unterricht im Hochschulbereich in der Regel weniger strukturiert ist als in anderen Bildungsstufen, sollten die Maßnahmen sorgfältig geplant und die Lernaktivitäten angemessen gestaltet werden, um

das Engagement und den Lernerfolg der Studierenden zu maximieren. Wirksames Feedback muss schnell und klar erfolgen, für die Studierenden und die jeweilige Aufgabe sinnvoll sein und Hinweise auf notwendige Verbesserungen geben. Es sollte auch darauf abzielen, mögliche Unzulänglichkeiten, das Abweichen vom geplanten Unterrichtsziel oder das Zurückfallen hinter den Zeitplan im Zusammenhang mit Prokrastination oder anderen Selbstregulierungsproblemen zu korrigieren. Daher muss das Feedback strategisch geplant werden, um das Arbeitspensum des Lehrers zu bewältigen, rechtzeitig einzugreifen und Momente für die Selbstkorrektur der Studierenden einzubeziehen, indem die Lehrer ihr Feedback auf Schlüsselmomente der Aktivitäten konzentrieren. Es ist wichtig, den Zeitpunkt, den Ort und die Abfolge der Feedback-Ereignisse zu planen [74].

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

Die Auswirkungen von authentischem Unterricht oder Erfahrungslernen allein auf den Erwerb von CT-Fähigkeiten und -Dispositionen wurden im Rahmen der aktuellen Intervention nicht bewertet. Dennoch wurden in den Critical Thinking Blended Apprenticeships Curricula Fallstudien, reale/authentische Probleme und kritische Vorfälle in den verschiedenen implementierten Modulen eingesetzt, um die theoretischen und praktischen Aspekte der vermittelten Themen miteinander zu verweben. Die Verknüpfung von Fakten- und Konzeptwissen, das sich die Studierenden aneignen müssen, mit praktischen Beispielen aus dem Berufsalltag erhöht die Lernmotivation der Studierenden und macht sie zu positiven Erfahrungen [65, 75], bei denen die Studierenden die Möglichkeit haben, verschiedene Ansätze zu erproben, um Probleme (selbständig) zu lösen. Daher sollten die Interventionen herausfordernd und nah an der Realität des Berufs sein und den Studierenden eine gewisse Entscheidungsfreiheit ermöglichen.

CTBACs müssen es den Studierenden ermöglichen, CT-bezogene Dispositionen zu entwickeln, die sie als nützlich für die beruflichen Ziele ansehen, die sie für sich selbst festlegen. Durch die Einbeziehung von Arbeitsmarkterfahrungen in die Ausbildung (entweder während der

Lehrzeit oder im Unterricht) werden unserer Ansicht nach CT-Fähigkeiten und CT-Dispositionen bei den Studierenden gefördert und verstärkt. Darüber hinaus wird davon ausgegangen, dass die Auswirkungen des authentischen und erfahrungsbasierten Lernens maximiert werden, sobald die Lehramtsstudenten in die Ausbildung eintreten und die Fähigkeiten und Dispositionen in realen Kontexten umsetzen.

## 5. Angemessenes Risiko eingehen/akzeptieren

Das CT setzt komplexe Probleme mit unsicheren Lösungen voraus. Nur in dieser Art von Situation können die Studierenden, die Auszubildenden und sogar die Ausbilder nicht auf frühere Erfahrungen zurückgreifen, um das Problem zu lösen. Wer in einer bestimmten Situation sein Gedächtnis einsetzt, kann seine Fähigkeiten zum kritischen Denken nicht nutzen, um das Problem zu lösen. Daher müssen die Studierenden in Situationen gebracht werden, in denen sie mehrere verschiedene Lösungen für ein Problem in Betracht ziehen und diejenige auswählen müssen, die sie für die geeignetste halten. Und vielleicht liegen sie dabei falsch. Folglich müssen sie immer wieder von vorne anfangen. Aber diese Möglichkeit des Scheiterns ist absolut notwendig, um CT-Fähigkeiten und CT-Dispositionen zu entwickeln. Deshalb muss bei der Planung der Ausbildungsprogramme der Fehler vorausgesehen und das Risiko des Scheiterns in Kauf genommen werden.

Das Begehen von Fehlern (verstanden als falsche Entscheidungsfindung), gefolgt von korrigierendem Feedback und dem Umgang mit Misserfolgen ist eine starke Lernerfahrung [76, 77]. Das korrigierende Feedback muss aus der Analyse der zugrundeliegenden Ursache resultieren, damit es konsequent ist [76].

Das Erleben und Erkennen von Fehlern fördert die Selbstregulierung (d. h. die Selbstbeobachtung und Selbsteinschätzung), die Nutzung alternativer Lösungen und die Metakognition. Darüber hinaus werden der spätere Gedächtnisabruf und die Fähigkeit, subtile Hinweise auf das Gesamtzenario richtig zu berücksichtigen, verbessert, während die

Auswirkungen von Entscheidungen mit hohem und niedrigem Vertrauen gemildert werden [76].

Die Erfahrung einer mangelhaften Entscheidungsfindung im Unterrichtskontext stimuliert die Entwicklung von CT-Fähigkeiten und CT-Dispositionen bei den Studierenden und wird erwartungsgemäß auftreten. Daher sollten Überwachungsstrategien und Feedback-Zyklen geplant und in die Interventionsstrategien integriert werden, um die Gewinne in CTBACs-Kursen zu steigern.

## 6. Reflektieren Sie die Veränderungen der CT-Fähigkeiten und CT-Dispositionen

Als Teil des Feedbacks oder parallel dazu ist die Reflexion über den Lernprozess ein integraler Bestandteil der Entwicklung von CT-Fähigkeiten und CT-Dispositionen und ist eine gefragte Haltung sowohl beim Lernen als auch im beruflichen Umfeld. Den Lernenden die Möglichkeit zu geben, über ihren Lernprozess zu reflektieren, ist eine Form, in der sie über die bloße Teilnahme hinausgehen und sich auf das "Wie" der Erledigung ihrer Aufgabe oder Aktivität konzentrieren können, indem sie erforschen, warum sie es tun [78]. Metakognition, eine höhere CT-Fähigkeit, beinhaltet das Bewusstsein für das eigene Denken oder das Denken über das Denken. Wie Nappi feststellt [79], ist Metakognition eine wesentliche Fähigkeit, die verfeinert werden muss, um zu erkennen, wie man lernt.

In der aktuellen Studie boten die Critical Thinking Blended Apprenticeships Curricula verschiedene Möglichkeiten zur Reflexion. Damit diese Möglichkeiten sinnvoll sind, sollten sie explizit sein. Wir setzen uns dafür ein, dass den Studierenden die Möglichkeit geboten wird, über die Vorgehensweise und die Endergebnisse der von ihnen geübten Aufgaben zu reflektieren. Dies führt dazu, dass die Studierenden die Stärken und Schwächen ihrer Leistung überdenken und bewerten und ein tieferes Verständnis dafür entwickeln, was von ihnen erwartet wird und ob sie die Aufgaben gemäß den geforderten Standards erfüllt haben. Dieser Prozess wird die Studierende auch in die Anwendung von CT-Fähigkeiten einbinden und ihnen



zeigen, dass die Aufgaben nicht nur das Abrufen von Informationen erfordern, sondern sie vielmehr dazu auffordern, diese zu analysieren, anzuwenden und neue Formen von Wissen zu schaffen. Wenn man den Studierenden beibringt, über ihr Denken nachzudenken (Metakognition), können sie zu einem tieferen Verständnis gelangen. Darüber hinaus können die Reflexion und Selbsteinschätzung ein angeleiteter Prozess sein, in den auch die Mentoren ausdrücklich einbezogen werden können. Gleichzeitig kann die Reflexion unter Gleichaltrigen auch durch Fokusgruppendifkussionen mit einem Fragenkatalog erfolgen, bei denen die Studierenden ihr Lehrmaterial und ihre Tagebücher austauschen und in entspannter Weise diskutieren können. Dieser Weg kann wahrscheinlich neue Möglichkeiten für die Entwicklung von CT eröffnen, da belastende Faktoren wie die akademische Leistung ausgeschlossen sind. Dennoch kann diese Art der Gruppenreflexion als Pflichtaufgabe organisiert werden und Blogs sowie Moodle können zu diesem Zweck genutzt werden.



## Literatur

[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 Sciences* 2022.

[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, Merfeldaitė O, Penkauskienė D, Pivorienė J, Railienė 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, Lithoxidou 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.

- [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.

- [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.

- [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.

- [53] Pnevmatikos D, Christodoulou P, Lithoxidou A, Georgiadou T. Designing Critical Thinking Blended Apprenticeships Curricula to Promote Reflective Thinking in Higher Education. 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. 316-28.
- [54] 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.
- [55] Lampert N. Critical Thinking dispositions as an outcome of undergraduate education. The Journal of General Education. 2007;56:17-33.
- [56] Bloch J, Spataro SE. Cultivating Critical-Thinking Dispositions Throughout the Business Curriculum. Business and Professional Communication Quarterly. 2014;77:249-65.
- [57] Marin LM, Halpern DF. Pedagogy for developing critical thinking in adolescents: Explicit instruction produces greatest gains. Thinking Skills and Creativity. 2011;6:1-13.
- [58] 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.
- [59] Europe Co. Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume. Strasbourg: Council of Europe Publishing; 2020.
- [60] Piccardo E, North B. The Action-oriented Approach: Multilingual Matters; 2019.
- [61] Kriaučiūnienė R, Targamadžė 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.
- [62] Colucciello ML. Relationships between critical thinking dispositions and learning styles. Journal of Professional Nursing. 1999;15:294-301.
- [63] 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.
- [64] 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.

[65] 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.

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

[67] 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.

[68] OECD. *Getting Skills Right: Skills for Jobs Indicators 2017*.

[69] 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.

[70] 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.

[71] 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.

[72] 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.

[73] 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.

[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 Psychology 2017. p. 8.

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



## Ergänzendes Material

**Ergänzende Tabelle 1.** Übersetzung der Fragebögen CTSAS-SF und SENCTDS in lokale Sprachen mit Hyperlinks.

### Fragebögen CTSAS-SF & SENCTDS in locale Sprachen

[CTSAS-SF & SENCTDS Scales in Griechisch](#)

[CTSAS-SF & SENCTDS Scales in Deutsch](#)

[CTSAS-SF & SENCTDS Scales in Portugiesisch](#)

[CTSAS-SF & SENCTDS Scales in Litauisch](#)

[CTSAS-SF & SENCTDS Scales in Rumänisch](#)

**Ergänzende Tabelle 2.** Deskriptive Statistik der CTSAS-SF-Items.

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
20. I figure out the assumptions implicit in the author's reasoning.	3.73	1.275	-.436	-.032	0.828	0.500

Items	Mean	Sd.	Skew.	Kurt.	K-S test	p
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
37. I use both deductive and inductive reasoning to interpret information.	4.00	1.330	-.419	-.259	0.766	0.600

Items	Mean	Sd.	Skew.	Kurt.	K-S test	p
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
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

Items	Mean	Sd.	Skew.	Kurt.	K-S test	p
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

**Ergänzende Tabelle 3.** Deskriptive Statistik der SENCTDS-Items.

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	<b>.047</b>
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	<b>.001</b>
13. I take notes so I can organize my thoughts.	5.19	1.653	-.900	.046	1.891	<b>.002</b>
14. I make simple charts, diagrams or tables to help me organize large amounts of information.	4.58	1.795	-.479	-.775	1.598	<b>.012</b>
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
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

Items	Mean	Sd.	Skew.	Kurt.	K-S test	p
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

**Ergänzende Tabelle 4:** Ladungen der Items in CTBACS\_SF

<i>Item</i>	<i>Interpretation</i>	<i>Analyse</i>	<i>Bewertung</i>	<i>Schlussfolgerung</i>	<i>Erläuterung</i>	<i>Selbstregulierung</i>
<i>1. I try to figure out the content of the problem.</i>	0.662					
<i>2. I classify data using a framework.</i>	0.661					
<i>3. I break the complex ideas into manageable sub-ideas.</i>	0.633					
<i>4. I observe the facial expression people use in a given situation</i>	0.386					
<i>5. I examine the values rooted in the information presented.</i>	0.654					
<i>6. I restate another person's statements to clarify the meaning.</i>	0.499					
<i>7. I figure out an example which explains the concept /opinion.</i>	0.594					
<i>8. I clarify my thoughts by explaining to someone else.</i>	0.422					
<i>9. I seek clarification of the meanings of another's opinion or points of view.</i>	0.536					
<i>10. I examine the similarities and differences among the opinions posed for a given problem.</i>		0.614				
<i>11. I examine the interrelationships among concepts or opinions posed.</i>		0.734				
<i>12. I look for supporting reasons when examining opinions.</i>		0.671				
<i>13. I look for relevant information to answer the question at issue.</i>		0.650				
<i>14. I examine the proposals for solving a given problem.</i>		0.701				
<i>15. I ask questions in order to seek evidence to support or refute the author's claim.</i>		0.666				
<i>16. I figure out if author's arguments include both for and against the claim.</i>		0.670				
<i>17. I figure out unstated assumptions in one's reasoning for a claim.</i>		0.619				
<i>18. I look for the overall structure of the argument.</i>		0.707				

<i>Item</i>	<i>Interpretation</i>	<i>Analyse</i>	<i>Bewertung</i>	<i>Schlussfolgerung</i>	<i>Erläuterung</i>	<i>Selbstregulierung</i>
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		
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		

Item	Interpretation	Analyse	Bewertung	Schlussfolgerung	Erläuterung	Selbstregulierung
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	
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

<i>Item</i>	<i>Interpretation</i>	<i>Analyse</i>	<i>Bewertung</i>	<i>Schlussfolgerung</i>	<i>Erläuterung</i>	<i>Selbstregulierung</i>
<i>54. I examine my values, thoughts / beliefs based on reasons and evidence.</i>						0.756
<i>55. I continuously assess my targets and work towards achieving them.</i>						0.673
<i>56. I review my reasons and reasoning process in coming to a given conclusion.</i>						0.728
<i>57. I analyse areas of consistencies and inconsistencies in my thinking.</i>						0.737
<i>58. I willingly revise my work to correct my opinions and beliefs.</i>						0.750
<i>59. I continually revise and rethink strategies to improve my thinking.</i>						0.786
<i>60. I reflect on my thinking to improve the quality of my judgment.</i>						0.763

**Ergänzende Tabelle 5:** Ladungen der Items in SENCTDS

<i>Item</i>	<i>Reflektion</i>	<i>Achtsamkeit</i>	<i>Aufgeschloss enheit</i>	<i>Organisation</i>	<i>Durchhalteve rmögen</i>	<i>Intrinsische Motivation</i>
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			



<i>Item</i>	<i>Reflektion</i>	<i>Achtsamkeit</i>	<i>Aufgeschlossenheit</i>	<i>Organisation</i>	<i>Durchhaltevermögen</i>	<i>Intrinsische Motivation</i>
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

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



## 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. We also want to thank the External Evaluation and Quality Committee/Steering Board, Caroline Dominguez (University of Trás-os-Montes and Alto Douro) for her review on the document.

© THINK4JOBS 2023