FACULTEIT PSYCHOLOGIE EN
PEDAGOGISCHE WETENSCHAPPEN

EDUCATION AND TRAINING RESEARCH UNIT

CENTRE FOR INSTRUCTIONAL PSYCHOLOGY AND TECHNOLOGY

Research Integration in Higher Education.
Prevalence and Relationship with Critical Thinking

An Verburgh

Proefschrift aangeboden tot het verkrijgen van de graad van Doctor in de Pedagogische Wetenschappen

Promotor: Prof. dr. J. Elen

2013

Research integration in higher education. Prevalence and relationship with critical thinking.

An Verburgh
Promotor: Prof. dr. Jan Elen

This doctoral dissertation focuses on the prevalence of research integration in higher education and its relationship with the development of critical thinking. Establishing a close(r) relation between teaching and research is considered as a dignified aspiration in higher education. The value of the link is defended on the idea that research integration is beneficial for student learning, such as students' development towards the ability to make reasoned judgements, in this doctoral dissertation referred to as critical thinking (CT). The recurrent proposals to make students more acquainted with the research field of their discipline, fit within this line of reasoning. By offering students more experience with research-based education or stated differently, by increasing research integration, the link between teaching and research would grow stronger and students would benefit.

However, empirical studies investigating the assumption that research integration influences the development of critical thinking, are lacking, except for studies on undergraduate research, a specific research integration approach. Consequently, the argument of a positive effect remains mainly conceptual and theoretical. Moreover, studies investigating the prevalence of research integration are limited. The frequency of research integration and its different manifestations is largely unknown.

The studies of this doctoral dissertation are done in Flemish higher education, with a main focus on university colleges. After the introduction (**Chapter 1**), that describes the background and the outline of the dissertation, an elaborated justification of the two main investigated variables on research integration is provided (**Chapter 2**). These variables are research-related learning outcomes (RR-learning outcomes) and research-integration approaches (RI-approaches). In **Chapter 3** the prevalence of RR-learning outcomes is studied within 45 programmes of different institutions of higher education. All module descriptions of these programmes were analysed to find indications for the pursuit of six RR-learning outcomes. Three distinct patterns in the prevalence of RR-learning outcomes between programmes were identified: a low-attention pattern with an overall low attention for the research-related goals; a results pattern with a higher attention for the research-related goals, especially for acquiring knowledge of research results (*Results*); and a critical-thinking pattern with the main focus on *Critical thinking*, *Practical research skills* and the *Competence to be a researcher*.

The subsequent studies concentrate on the first year of higher education. In **Chapter 4** RI-approaches in the first year of four bachelor's programmes at university colleges are documented. The results revealed differences in the prevalence of different variants of RI-approaches. The most frequent RI-approach was an approach in which students are confronted with research results as if they were facts. Involvement in authentic research was rare. In addition six module types were identified. Module types are specific combinations of one of more RI-approaches within one module. These module types appeared to have privileged relationships with specific RR-learning outcomes (**Chapter 5**). Acquiring knowledge of research results (*Results*) or *Competence to be a researcher*, does not differentiate between first year module types, while s*pecifications of practical research skills*, *Critical thinking* or the attention for underpinnings of research (*Underpinnings*) does. **Chapter 6** deals with the measurement of the development of CT. Since no validated instrument to assess CT in Dutch was available, two tests were translated and validated. The study investigated the psychometric quality of a translation of the Cornell Critical Thinking Test (CCTT) and the Halpern Critical Thinking Assessment (HCTA) in a sample of Dutch-speaking first year students majoring in educational sciences. Results showed higher content validity and preference by students for the HCTA, while administration time for the CCTT was lower. Neither test showed a high overall reliability and there are questions about the construct validity of both tests. Based on the results of this study the Scipio-test was developed and used in the following study (**Chapter 7)** in which the relation between research integration and CT was studied. Results showed that students develop in CT. However, although students from different programmes were confronted with different research integration practices, they did not differ in their CT development.

The doctoral dissertation ends with a discussions of the findings (**Chapter 8**). Whilst the results are coloured by the local context of higher education in Flanders, the findings are also relevant for the larger field as it challenges the assumption of a simplistic relationship between research integration and student learning. The broad interpretation of research integration used in this doctoral dissertation provokes the field to use perspicuous language when discussing research integration and the research-teaching nexus. In addition, the results point at the value of fine-grained analyses of research integration practices at the right level of specification for improving our understanding of the complexities involved. An agenda for future research on research integration concludes the dissertation.

Onderzoeksintegratie in het hoger onderwijs: Frequentie en relatie met kritisch denken.
An Verburgh
Promotor: Prof. dr. Jan Elen

Dit proefschrift richt zich op de mate waarin onderzoeksintegratie in het hoger onderwijs voorkomt en het verband tussen dat voorkomen en de ontwikkeling van kritisch denken. In het hoger onderwijs wordt het creëren van een hechte(re) band tussen onderwijs en onderzoek als waardevol aanzien. Dit idee is gestoeld op het argument dat onderzoeksintegratie bevorderlijk is voor wat studenten leren, zoals het ontwikkelen van kritisch denken. In dit proefschrift wordt kritisch denken beschouwd als de bekwaamheid om beredeneerd te oordelen. De voorstellen om studenten meer ervaring met onderzoek uit hun discipline te laten opdoen of de vragen naar onderzoeksgebaseerd onderwijs sluiten aan bij deze argumentatielijn. Door meer onderzoeksintegratie in het onderwijs zou het verband tussen onderzoek en onderwijs sterker worden en zouden studenten meer leren. Er is echter geen empirisch onderzoek naar de juistheid van de argumentatie dat onderzoeksintegratie kritisch denken positief beïnvloedt. Een uitzondering hierop zijn studies naar het effect van studentenonderzoek (undergraduate research), een specifieke aanpak van onderzoeksintegratie waarin de student zelf authentiek onderzoek doet onder begeleiding van een onderzoeker. Doordat onderzoek naar het effect van onderzoeksintegratie ontbreekt, is blijft de argumentatie conceptueel en theoretisch. Daarenboven is er slechts een beperkt aantal studies naar de mate waarin onderzoeksintegratie voorkomt. De frequentie van verschillende uitingen van onderzoeksintegatie is grotendeels onbekend.

Dit proefschrift is uitgevoerd in het hoger onderwijs in Vlaanderen, voornamelijk in hogescholen. **Hoofdstuk 1** geeft de achtergrond en opbouw van het proefschrift weer. In **Hoofdstuk 2** wordt uitvoerig ingegaan op de invulling van de twee onderzochte hoofdvariabelen van onderzoeksintegratie. Dit zijn onderzoeksgerelateerde onderwijsdoelen en aanpakken van onderzoeksintegratie. Voor de studie beschreven in **Hoofdstuk 3** werd bij 45 programma's nagegaan hoe vaak onderzoeksgerelateerde onderwijsdoelen voorkwamen. Hiervoor werden de vakbeschrijvingen van alle opleidingsonderdelen van die opleidingen bestudeerd op indicaties over het nastreven van zes onderzoeksgerelateerde onderwijsdoelen. De onderzochte opleidingen passen elk binnen een van de drie gevonden patronen: een patroon met een beperkte aandacht voor onderzoeksgerelateerde doelen; een patroon met vooral aandacht voor het verwerven van onderzoeksresultaten en een derde patroon, het kritisch denkenpatroon, met een uitgesproken aandacht voor kritisch denken, praktische onderzoeksvaardigheden en de competentie om een onderzoeker te zijn.

De daaropvolgende studies concentreren zich op het eerste jaar in het hoger onderwijs. Hoofdstuk 4 brengt de aanpakken van onderzoeksintegratie van vier hogeschool bachelorprogramma's in kaart. De resultaten wezen op verschillen in de frequentie van elk van de aanpakken. De meest frequente aanpak was die waarin studenten in contact komen met onderzoeksresultaten als feiten. Betrokkenheid bij authentiek onderzoek was zeldzaam. Daarnaast werden zes types van opleidingsonderdelen onderscheiden. Dat zijn specifieke combinaties van één of meer aanpakken van onderzoeksintegratie binnen hetzelfde opleidingsonderdeel. Deze moduletypes bleken samen te gaan met bepaalde onderzoeksgerelateerde doelen (**Hoofdstuk 5**). De onderzoeksgerelateerde doelen verwerven van kennis van onderzoeksresultaten (*Resultaten*) en de *competentie om onderzoeker te zijn* verschilden niet naargelang het type maar bepaalde *specifieke praktische onderzoeksvaardigheden*, *kritisch denken* en de aandacht voor de theoretische en methodologische onderbouw van onderzoek *(Onderbouw)* deden dit wel. Omdat geen gevalideerd Nederlandstalig instrument voor kritisch denken beschikbaar was, werden in **Hoofdstuk 6** twee testen vertaald en bij eerstejaarsstudenten Pedagogische wetenschappen gevalideerd: de Cornell Critical Thinking Test (CCTT) en de Halpern Critical Thinking Assessment (HCTA). De resultaten toonden een hogere inhoudsvaliditeit en voorkeur van studenten voor de HCTA, terwijl de administratietijd voor de CCTT lager was. Geen van de testen toonde een hoge betrouwbaarheid en constructievaliditeit. Op basis van de resultaten werd de Scipio-test ontwikkeld en in de volgende studie gebruikt (**Hoofdstuk 7**). In deze studie werd de relatie tussen onderzoeksintegratie en kritisch denken onderzocht. De resultaten toonden aan dat studenten groeiden in kritisch denken. Ondanks verschillen tussen de programma's in onderzoeksintegratie, werd er tussen de programma's geen verschil in groei in kritisch denken vastgesteld.

Het proefschrift eindigt met een discussie over de bevindingen en een onderzoeksagenda (**Hoofdstuk 8**). Hoewel de resultaten gekleurd zijn door de context van het Vlaamse hoger onderwijs, zijn de resultaten ook waardevol voor een groter publiek omdat ze de veronderstelling van een eenvoudige relatie tussen onderzoeksintegratie en het leren van studenten in vraag stellen. De brede invulling van wat geldt als onderzoeksintegratie daagt uit tot heel precies taalgebruik bij het bespreken van onderzoeksintegratie en de relatie tussen onderwijs en onderzoek. De resultaten wijzen op de waarde van fijnmazige analyses op het juiste abstractieniveau om de complexiteit van onderzoeksintegratie beter te vatten.

Word of thanks

Voor u ligt mijn doctoraat. Een mijlpaal is bereikt. De weg ernaartoe was soms kronkelig en langzaam, één enkele keer zelfs een heel scoringsproces achteruit maar het was vooral uitdagend, boeiend en leerzaam. Hoewel het mijn weg was, was ik nooit alleen onderweg. Ik had het geluk omringd te zijn door veel mensen, die elk op hun manier een bijdrage hebben geleverd aan dit proefschrift. Het feit dat ik deze mijlpaal heb bereikt, is dankzij hen!

Jan, bedankt om me de kans te geven om onder jouw begeleiding te doctoreren. Je leerde me het kaartlezen van onderzoek doen. Je liet me toe om onverharde wegen en zelfs zandpaden te verkennen. Bedankt voor deze vrijheid en het vertrouwen die je me schonk om mijn eigen weg te kiezen. Bedankt voor je toewijding en je enthousiasme zelfs voor kleine stappen. Je vragen hielpen me om mijn gedachten te verfijnen. Als ik een redering tracht te doorgronden, komt spontaan “Maar oei, pas op” in mij op. Je wist me steeds op het juiste moment de juiste suggesties en aanmoedigen te geven als ik dreigde van de weg te raken of als ik door het stof de weg niet meer zag. En je wist ook op welke momenten je me beter met rust liet. Bedankt om ervoor te zorgen dat ik in al mijn naïviteit mijn rugzak niet te zwaar belaadde.

Wendy, vanaf het begin van mijn assistentschap, hebben we samen het kluwen van onderzoeksintegratie proberen te ontwarren. Ik ben blij dat we een groot deel van onze reis samen hebben afgelegd. Samen hebben we voor IOO en IOO^2 half Vlaanderen afgereisd en ook een deel van de wereld gezien. Onze reisvoorbereidingen en -uitvoeringen waren dan wel niet altijd optimaal, ze leverden wel hilarische momenten op! Nu moet jij nog even doorzetten en dan kunnen we op reis naar Cadiz!

Sigrid, ook jij hebt een wezenlijke bijdrage geleverd aan dit proefschrift, waarvoor oprechte dank. Je vriendelijkheid en nauwgezetheid zijn een voorbeeld.

Dawit and Marie, you came in later but due to your critical presence the critical thinking track of the route gained momentum! I look forward seeing the results of your studies!

I would like to thank the members of my project commission, Mick Healey, Maarten Simons, Eddy Van Avermaet, Jan Van Driel en Piet Verhesschen. Their comments and suggestions helped me to keep on track. I would also like to thank the voluntary readers of my texts: Paula, Dawit, Marie, Wendy, Griet, Geraldine, Nick, Sigrid and Marleen. You all helped me refining my thoughts and text.

Diane Halpern and Heather Butler deserve also a special word of thanks. Thanks to Diane we were allowed to translate the HCTA and thanks to their meticulous answers on the nearly never-ending flow of questions of Sigrid and me, we were able to translate the test and grade the students' answers appropriately.

Zonder de bereidwillige medewerking van de docenten en studenten van de betrokken opleidingen was dit onderzoek onmogelijk geweest. Ik ben hen dan ook heel dankbaar voor hun tijd. In het bijzonder wil ik de betrokken opleidingshoofden bedanken voor hun steun aan mijn doctoraatsonderzoek. Dankzij jullie gingen de poorten van jullie opleidingen voor mij open! Jullie enthousiasme was zelfs zo groot dat er nog data liggen te wachten!

Rianne bedankt om me op weg te helpen met de IRT-analyses en je gedegen ondersteuning. De collega's van het CIP&T wil ook graag bedanken voor de leuke alma-, koffie en ijsjesmomenten. Ik geef toe dat ik er de laatste tijd niet zo vaak bij was maar vanaf nu komt daar terug verandering in. In het bijzonder wil ik Greet en Marleen bedanken. Greet, bedankt voor de leuke babbels, de betrokkenheid en alle momenten dat je ben ingesprongen om de kinderen op te vangen! Marleen, bedankt voor je rustgevende geruststellende aanwezigheid en voor je opbeurende woorden. Je positieve ingesteldheid is inspirerend. En uiteraard ook dankjewel voor al je hulp bij mijn werk. Zonder jou was ik nu wellicht nog aan het verbeteren!
Daarnaast een woordje van dank voor mijn collega-onderzoekers van OHOO, in het bijzonder Didi. Bedankt voor de vele discussies – op een of andere manier vooral tijdens de lunch voor onze vergaderingen - en de leuke en interessante momenten op congressen.

Daarnaast wil ik graag mijn vrienden en (schoon)familie bedanken. Bedankt voor jullie steun, voor jullie interesse in mijn werk (hoewel dit proefschrift wellicht niet jullie favoriete literatuur is), voor jullie volgehouden interesse (hoewel jullie je wellicht afvroegen waarom dat allemaal zo lang moest duren), voor jullie begrip (toen ik weer eens zei dat ik niet kon komen omdat ik wou schrijven of op congres was), voor de afleiding die jullie me bezorgden, voor de praktische hulp die jullie boden. Vanaf nu halen we alle gemiste en uitgestelde wandelingen, weekendjes, barbecues en spelletjes avonden in! Paul en Ria, ik waardeer jullie aanwezigheid enorm en hoop dat nog lang te kunnen doen. Moety en Padre, bedankt om me te leren dat de gemakkelijkste weg niet perse de beste of interessantste keuze is, en dat je vooral moet kiezen voor de weg die je interesseert. Met jullie als voorbeeld weet ik dat je met volharding veel kan bereiken. Bart en Greet, Mieke en Dimitri, Leen en Adrien, Erwin en Aukje, jullie hebben me allemaal, elk op jullie manier geholpen! Dankuwel hiervoor.

Nico, de afgelopen zeven jaar was dit werk er een wezenlijk onderdeel van mijn leven en dus ook van ons en jouw leven. Het krijgen van drie kinderen tijdens een doctoraat is misschien niet de meest handige gang van zaken. Maar dankzij jouw inzet was het wel haalbaar en ik had het voor geen geld willen missen. Jouw aandeel in dit proefschrift is onnoemelijk veel groter dan de lay-out en je excel-script! Ik ben je dankbaar dat ik samen met jou onderweg ben en dat we samen de weg van ons leven bewandelen. Bij jou kom ik thuis.

Roos, Kaat en Raf, dankjulliewel voor jullie vrolijkheid, jullie geduld en jullie ongeduld. Jullie maken mij ieder dag opnieuw duidelijk wat echt belangrijk is!

En voor we verder gaan, iemand een ijsje?

An Verburgh
Leuven,
Juli 2013

Table of Contents

Chapter 1: Introduction[1]

1.1 The Rationale in a Nutshell

The idea that all students – irrespective of their career prospects – may benefit from research is central to the recommendations to strengthen the role of research in the undergraduate curriculum (Brown & Mc Cartney, 1998; Heggen, Karseth, & Kyvik, 2010; Levy & Petrulis, 2012). Research integration would be beneficial for both learning specific research skills and for acquiring competences that transcend research. One of these competences is the development of critical thinking (CT) (Heggen et al., 2010; Lopatto, 2009). The development of CT is a central goal in higher education (King & Kitchener, 2004) as it is considered to be part of a research disposition (van der Rijst, 2009) and to be crucial to function in a complex contemporary society (Barnett, 1990; Bok, 2006).

Despite these firm claims, there is only limited empirical research about the effect of integrating disciplinary research into the curriculum on student learning, and consequently also on CT. Therefore the argument for more research integration remains largely conceptual and theoretical. Given the nature of the empirical studies on research integration into teaching, we lack insight in two fundamental issues: (1) the prevalence of research integration, and (2) the effect of research integration on student learning, especially on CT. This doctoral dissertation contributes to the empirical investigation of the research-teaching nexus by concentrating on these two fundamental issues.

[1] Partially based on Verburgh, A., Elen, J., & Lindblom-Ylläne, S. (2007). Investigating the myth of the relation between teaching and research in higher education: A review of empirical research. *Studies in Philosophy and Education, 26,* 449-465.

1.2 The Research-Teaching Nexus

Institutes of higher education are typically involved in research, teaching and public service (Rhodes, 2004). A fundamental issue in higher education is the relationship between research and teaching (Clark, 1997). Since research and teaching are two important activities within the same institution, and faculty are very often involved in both, the question about the relationship between the two is recurrent (Barnett, 2005).

The contemporary debate on the research-teaching nexus in higher education is influenced by recent changes in society and in the field of higher education. An important change is the shifting role of different institutes of higher education. Currently, in most European countries all types of higher education are expected to do research (Kyvik & Skodvin, 2003). This implies that all institutes of higher education, also those that were traditionally more focused on teaching, are also expected to be involved in research and to offer research-based education to their students (Kyvik, 2009; Kyvik & Lepori, 2010; Witte, van der Wende, & Huisman 2008). In addition, other developments affect the relationship between research and teaching too, such as: the move to mass higher education and its influence on the amount of time available for research and teaching; changes in the nature of research and teaching; changes in the policy context; shifts in the ideas about the nature of knowledge; the increased research expectations and the larger dependence on external research funding (Brew, 1999, 2003; Prince, Felder, & Brent, 2007; Simons & Elen, 2007).

Often, one tacitly considers the link between research and teaching as valuable (Prince et al., 2007). This frequently coincides with a normative position, which states that the link should be stronger than it currently is (Trowler & Wareham, 2008). Establishing a close(r) relationship between research and teaching is seen as a dignified aspiration. Brew (2013) relates a close link between research and teaching to "the academic project of the university" (p. 2), Robertson (2007) connects it to the "culture and mystique of the modern university" (p. 541). A close link fits with the Humboltian idea of the university as a place where research and teaching are united (Simons, 2006). Research and teaching would be interdependent, reinforce and enrich each other (Hughes & Tight, 1995). At the minimum, there would be a positive effect of research on teaching: research improves teaching, and teaching would not be as good if there were no research (Brown & Mc Cartney, 1998; Taylor, 2007; Prince et al., 2007). Closely related to the idea of a beneficial effect of research on teaching is the idea of a positive effect of research on student learning (e.g., Barnett, 2000; Brew & Boud, 1995; Neumann, 1992).

A repeated proposal to strengthen the link or to enhance student learning is to let students become more acquainted with research in the discipline of their study, by offering more research-based education both within the curriculum and as an extra-curricular activity (Boyer Commission, 1998; Brew, 2006, 2013; Heggen et al., 2010; Trowler & Wareham, 2008; Visser-Wijnveen, Van Driel, van der Rijst, Visser, & Verloop, 2012). The idea of research integration into teaching[2] echoes in higher education research and policy (Baxter Magolda, 2004; Brew, 2006; Heggen et al., 2010). The Boyer commission (1998) reproached American research universities for inadequately preparing undergraduate students for the challenges of professional life or graduate study and proposed to stimulate student learning through inquiry learning and research opportunities. Similarly, the OECD (1998) argued that all students in higher education would benefit from learning in a research and scholarly culture and that short inquiry projects to solve practical problems are a good way to introduce students to research and critical inquiry. The European University

[2] In this doctoral dissertation the expression "research integration into teaching" or "research integration" is used. This expression was preferred above research-based education or inquiry learning because some authors use it to refer to a specific kind of research integration (E.g., Healey, 2005b; Visser-Wijnveen, van Driel, van der Rijst, Visser, & Verloop, 2012; Zimbardi & Myatt, 2012).

Association (2007) considers research-based education as a strength of European universities and states that students need to gain research experience and develop research-related skills. These recommendations build on the idea that integration of research into teaching stimulates student learning. The Boyer commission (1998) states that "students would benefit from integrating their [teachers'] research and teaching experiences" (p. 8). More specifically, research integration helps in developing specific research skills and acquiring competences that transcend research, such as the development of CT (Heggen et al., 2010; Lopatto, 2009). Because the development of these competencies is highly valued in contemporary society (Brew, 2013; Casar, 2000; Durning & Jenkins, 2005; Jenkins, 2000), research integration is also valuable for students who do not envisage an academic research career. Therefore, a close (and positive) relationship between research and teaching is valuable for all students and should be aspired for.

However, at least three remarks can be made on this line of argumentation. A first remark relates to the inconclusive empirical evidence about the –often unstated– assumption of the tight link between research and teaching being positive (Brew & Boud, 1995; Visser-Wijnveen, 2009). Theoretically, different types and directions of relationships are possible. Theories and arguments are available for each of them (see Hattie & Marsh, 1996, for a detailed overview). First, the relationship between research and teaching may be a negative one. This fits with the incompatibility thesis, which conceives research and teaching as different and incompatible activities (Clark, 1997). Research might have a negative impact on teaching or, teaching might have an unfavourable impact on research. Second, research and teaching might be independent of one another. In that case research and teaching are unrelated. Third, research and teaching might be positively related. This compatibility thesis assumes research and teaching to be two compatible activities (Clark, 1997). The positive relationship can be mutual or unidirectional. In the former, they mutually reinforce one another, in the latter, research has a positive impact on teaching or the other way around. Hence, the viewpoint of a positive impact of research on teaching is only one possible way to look at the relationship.

Empirical research on the relationship between research and teaching is inconclusive (Brew & Boud, 1995; Visser-Wijnveen, 2009). On the one hand, correlation studies generally try to find a relationship between research productivity and teaching quality. A seminal review of Hattie and Marsh (1996) revealed a slightly positive overall correlation between teaching effectiveness and research output, but the correlation was close to zero (r = .06). Allen (1996) and Braxton (1996) drew similar conclusions. Other aspects of the relationship, such as time allocation, were also studied (Colbeck, 1998; Hattie & Marsh, 1996; Marsh & Hattie, 2002). Time spent on research was shown to be independent of teaching quality (Hattie & Marsh, 1996).

On the other hand, experience-related studies focus on perceptions of teachers[3], administrators or students on the relationship between research and teaching and how they conceive possible advantages and disadvantages of that relationship. These studies reveal a strong conviction of a positive effect of research on teaching (Durning & Jenkins, 2005; Kreber, 2000; Leisyte, Enders, & de Boer, 2009; Neumann, 1992, 1994; Robertson & Bond, 2001; Rowland, 1996; Zamorski, 2002). However, in these studies respondents also gave examples of a negative or absent effect, such as an unbalanced curriculum and teachers' limited availability for students since teachers might be more directed by their personal research interests than by students' learning.

Thus, the empirical evidence on the link between research and teaching does not equivocally point in the direction of a positive relationship. Prince and his colleagues (2007) attribute these conflicting results to two different underlying propositions: (1) correlation studies investigate whether research does support

[3] In this doctoral research 'teacher' will be used to refer to the person who teaches and abstraction is made of the other duties, responsibilities or background of the teacher.

teaching in practice, and (2) experience related studies investigate whether research can support teaching.

A second remark on the argumentation in favour of a close link, is that it rarely specifies the precise meaning of "a close link", its elements and the processes involved (Spronken-Smith, Walker, Batchelor, O'Steen, & Angelo, 2012; Trowler & Wareham, 2008). Does "research" refer to the research activity of the teacher, the confrontation of students with research, the research conceptions or something else? Is "teaching" understood as the overall teaching approach of teachers, or rather as their approaches concerning research-related content and skills? Does it concern teaching skills or students' appreciation? And which aspects of learning are meant?

When "research" refers to the research activity of the teacher, different interpretations may prevail. It might refer to the research excellence of the teacher; in that case, one could expect that the better the teacher is as a researcher (the more articles she produces, the more research awards she has received, ...), the better the teaching would be. Alternatively, it might refer to the teacher's research experience. Within this interpretation a teacher with research experience is assumed to teach better than a teacher without such experience but that it is not necessary for a teacher to be an active researcher. When "research" refers to the confrontation of students with research, it could be labelled as "research integration into teaching" and likewise refer to a set of very different practices. Some possibilities are that students have to do research themselves, that teachers use research results in their courses or that students learn about research methods. The assumption to be made here is that the more students have to do research themselves, the more they appreciate their learning environment. Likewise, one could assume that the more students are confronted with cutting edge research results, the better their understanding of the research domain would be.

Moreover, the teaching side of the research-teaching nexus could also refer to "research integration". In that case –still under the assumption of a positive effect of research on teaching– a dependency of research integration on the teacher's research activity is expected. For example: the better the teacher is as a researcher, the more up-to-date content she uses in her teaching or the better she teaches the students research methods. Another assumption could be that teachers with research experience give students more research-related assignments than teachers without research experience. In addition, teaching could also refer to the overall quality of teaching or students' appreciation for their learning environment. "Learning" could be understood as learning in relation to research, such as acquiring research skills or developing a research attitude, but also as learning without this relationship to research, such as the acquisition of professional competencies or personal development. A possible assumption in this respect might be that students, who work on a project under the supervision of a highly productive researcher, develop more research skills than students who work for a less productive researcher.

This exemplary listing of possible interpretations and assumptions highlights the complexity of the relationship between research and research and teaching integration. It should also be noted that all examples take the researcher as starting point. The department or the type of institution could be considered as well, rendering the issue even more complex. An assumption at the level of the department could be that the higher the ranking of the research group is, the better the teaching becomes. A possible assumption at the level of the institute might be that students in research-intensive institutions gain a better understanding of the discipline than students in teaching-intensive institutions. It is thus necessary to clarify the meaning of each element in the research-teaching discussion in order to avoid confusion and talking at cross purposes.

A third remark concerns the limited empirical research on research integration and its effects on student learning (Verburgh, Elen, & Lindblom-Ylänne, 2007). Hattie and Marsh (1996) concluded in their review that the frequency of research integration remains to be investigated. And hitherto, it is still largely unknown how frequently students come in contact with research during their studies. The number of publications describing good practices concerning research integration is abundant (e.g., Farrand-Zimbardi, van der Burg, & Myatt, 2010; Gunn, Draper, & Mckendrick, 2008; Healey, Lannin, Stibbe, & Derounian, 2013; Jenkins, Healey, & Zetter, 2007; Malcolm, 2008; Zubrick, Reid, & Rossiter, 2001). These publications indeed highlight actual teaching practices and specify modules or activities in which students are confronted with research. However, they do not answer the question about the prevalence of research integration in the curriculum. They do not reveal how common the selected "good practices" are and consequently give no indication on how often students are confronted with research integration during their programme.

Studies on the role of research within the curriculum are limited. Sin (2012) investigated the role of research within the master's curriculum and found that the development of research skills was generally considered as being important, while there were major differences between programmes. Van der Rijst and Jacobi (2010) scanned three bachelor's programmes of the same university on their research integration practices. They found that students were very often involved in research-like activities. However, students seldom did authentic research, i.e. research oriented to the development of knowledge and insights new to the discipline. In addition, programmes seemed to continuously seek a balance between acquiring knowledge from research results and mastering the process of conducting research.

Moreover, research that attempts to assess the expected learning benefits of research integration is largely lacking. This is surprising, given the centrality of the claimed benefits of research integration on student learning. One exception is the evaluation studies on students' research participation programmes (often referred to as undergraduate research, Lopatto, 2009). Research participation programmes offer students opportunities to be involved in research. They predominantly imply that students work on an authentic research project under close supervision of a researcher during summer or during a semester (Lopatto, 2009). Admission to these programmes is often competitive or restricted to advanced or higher achieving students (Hathaway, 2002; Hunter, Laursen, & Seymour, 2007; Seymour, Hunter, Laursen, & Deantoni, 2004). These studies investigate effects on student learning, among other outcomes such as student attrition, or continuation for advanced studies. The studies show positive effects on student learning, such as the development of research skills, personal growth and development of CT (e.g., Adhikari & Nolan, 2002; Bauer & Bennett, 2003; Hunter, Laursen, & Seymour, 2007; Kardash, 2000; Rauckhorst, Czaja, & Baxter Magolda, 2001; Seymour, Hunter, Laursen, & Deantoni, 2004). However often the design of these evaluation studies are limited and control groups often lack (Seymour et al., 2004). When controlling for different background variables effects seem to diminish (e.g., Carter, Ro, Alcott, & Lattuca, 2013).

Overall, the above discussion indicates that there is the idea of research integration being beneficial for students and that as a consequence the role of research in the curriculum should be strengthened. Nonetheless, the frequency of research integration and its relationship to student learning is seldom investigated.

Insight into the prevalence of research integration could strengthen our understanding of the research-teaching nexus. We currently do not know whether students' confrontation with research occurs on a daily basis, is limited to a final year research project or whether it is something in between. It is unknown which practices of research integration are more common than others. If research integration should be strengthened, it is essential to have an indication of the current situation of research integration in the curriculum.

Insight into the effects of research integration on student learning would help to underpin the argumentation in favour or against research integration. This would provide valuable information in order to develop evidence-based policies. Currently it is unknown whether student development depends on the frequency of research integration and if different research integration practices have a differential effect. The perception studies glance at the effects of research on teaching and learning by demonstrating the existence of this conviction. However, the existence of a conviction about an effect does not automatically imply the effect itself (as demonstrated in the correlation studies), nor does it reveal how often research is integrated. Evaluation studies on the effect of research participation on student learning are only partially answering the question, since they do not investigate the effects of the day-to-day experiences of (all) students with research, but merely the effects of a very specific learning experience of (a selected group of) students instead.

This doctoral dissertation will therefore focus on the following two questions: (1) What is the prevalence of research integration, and (2) What is the effect of research integration on student learning. Since student learning is too broad as a concept and the development of critical thinking (CT) is mentioned as a positive effect of research integration that transcends research, the focus will be on CT.

1.3 Student Learning: Critical Thinking

CT is indispensible in contemporary society (Barnett, 1990; Bok, 2006; Halpern, 2003). Because of the increasing complexity of the society and the wealth of information available in it, it is important to be able to make reasoned judgements about ill-structured problems and to interpret and judge information in an appropriate way (Halpern, 2003; King & Kitchener, 1994). Moreover, CT is considered essential for active and engaged citizenship (Kuhn, 1999; Moore, 2004). Therefore it is no surprise that CT is a central intended learning outcome in higher education (Halpern, 1998; King & Kitchener, 2004; Tapper, 2004; Tsui, 2001). Governments, accreditation agencies and employers long for alumni who are able to think critically (Harvey, 2005; Pithers & Soden, 2000) and universities claim students develop in CT during their studies (Davies, 2011).

CT is a multilayered concept with many different definitions (Bailin, Case, Coombs, & Daniels, 1999) and connected with different concepts like reflective thinking (King & Kitchener, 2004), epistemological beliefs (Schommer & Walker, 1995), or epistemological reflection (Baxter Magolda, 2004). Nevertheless, all definitions specify that CT consists of a set of skills, such as the skill to analyse arguments, identify key elements, synthesise information, or draw conclusions (Erwin, 2000; Halpern, 1999; Pascarella & Terenzini, 2005). Moreover, the majority of researchers indicate that, in addition to this set of skills, CT includes a dispositional component (Angeli & Valanides, 2009; Giancarlo & Facione, 2001; Ku & Ho, 2010; Pascarella & Terenzini, 2005). It refers to a motivational dimension of CT (Halpern, 2003). The dispositional component includes open mindedness, truth-seeking and inquisitiveness (Facione, 2010; Giancarlo & Facione, 2001).

Finally, most researchers agree that CT is domain-general to a certain extent, implying that CT, mastered in one domain, can also be used and applied in a wider range of domains (Angeli & Valanides, 2009).

The development of CT can be stimulated through education (Baxter-Magolda , 2004; Halpern, 2003; Niu et al., 2013; Pascarella & Terenzini, 2005; Terenzini, Springer, Pascarella, & Nora, 1996; Tsui, 1999, 2002). Research integration is claimed to have a positive effect on CT development. And there are positive empirical indications of the effect of undergraduate research on CT (Kuh, 2007; Seymour et al., 2004). In addition, there are different suggestions about instructional approaches that could effectively develop students' CT performance (Abrami et al., 2008; Tiruneh, Verburgh, & Elen, 2013; Tsui, 1999; 2002). For example, integrating CT instruction explicitly within subject-matter content has been shown to be more effective in developing CT than an immersion approach (e.g., Abrami et al., 2008). In an immersion approach, CT principles remain implicit, while it also aims at encouraging students to reason along CT principles (Ennis, 1989). Moreover, teacher modelling has positive effects (Halpern, 1998). Explicitly targeting epistemological assumptions and research methods are also shown to positively affect CT (Hofer, 2004; Kronholm, 1996). Hofer (2004) suggests that "students need instruction not only in *what* psychologists and chemists know, but also in *how* they know what they know" (p.

161) and she advises teachers to pay more attention to research methods and principles of justification in a field of knowledge. Kronholm (1996) found that carefully led discussion about epistemological assumptions established an improvement of epistemic reasoning, compared to control group students. The use of ill-structured, messy, complex problems can also serve as a trigger for applying and developing CT (Halpern, 1998; King & Kitchener, 2004). In addition, critiquing a paper, class presentation and discussions, writing and rewriting with much attention to analysis, synthesis, and refinement of ideas and collaborative work have all proven their effectiveness in developing CT (Astin, 1993; Tapper, 2004; Tsui, 1999; 2002). Some of these teaching practices are likely to occur when research is integrated into the curriculum.

However, despite the possibility of education to stimulate the development of CT, the success of higher education to actually do so is limited. Studies on CT development in higher education revealed that students indeed grow in CT, but only to a limited extent and mainly during the first (two) year(s) (Arum & Roksa, 2011; Astin, 1993; Bers, McGowan, & Rubin, 1996; Evens, Verburgh, & Elen, 2013; Giancarlo & Facione, 2001; Hagedorn et al., 1999; Miller, 1992; Saucier, 1995; Pascarella, Blaich, Martin, & Hanson, 2011). Arum and Roksa (2011), for example, found in their study that on average there was an increase of only .18 standard deviations during the first two years of higher education and almost half of the students showed no growth at all on CT.

These findings elicit the question if research integration indeed adds to the development of CT skills. Although there are indications of positive effects of undergraduate research, it is unknown if this also applies to a broader conceptualisation of research integration. It is unknown whether students, who are more confronted with research integration, show more development in CT than students who are less confronted with it, nor is it known whether different research integration practices have a differential effect.

1.4 Conceptual Framework

The design of this doctoral dissertation is inspired by curriculum studies, although this type of study is not common in the literature on the research-teaching nexus (see Carter et al., 2013, for an exception). Curriculum studies on CT investigate the frequency of use of diverse teaching approaches and characteristics of the institutions and programmes. They look for similarities and differences between these variables and relate these to the development of CT. Tsui (1999, 2002) for example investigated whether differences in educational approaches between curricula could explain differences in the development of CT. Seifert, Pascarella, Colangelo, and Assouline (2007) studied whether differences in teaching approaches between honour programmes and standard programmes could account for differences in CT development.

The design of this doctoral dissertation can be situated within the conceptual framework of Terenzini and Reason (2005; 2013). Based on 35 years of research (Pascarella & Terenzini, 1991; 2005), Terenzini and Reason (2005; 2013) propose a conceptual framework to analyse influencing factors on student learning (and persistence) in the first year and beyond. The model focusses on identifying factors in the college experience that influence students' success and that are somewhat controllable by faculty members and policy makers. According to the model, students enter higher education with a set of characteristics and experiences, such as personal experiences and academic preparation. The model hypothesises that these entrance characteristics have an impact on the individual curricular, classroom and out of class experiences of students in higher education, which are influenced by their peer environment to some extent. These individual student experiences subsequently shape the learning experiences and learning outcomes of students.

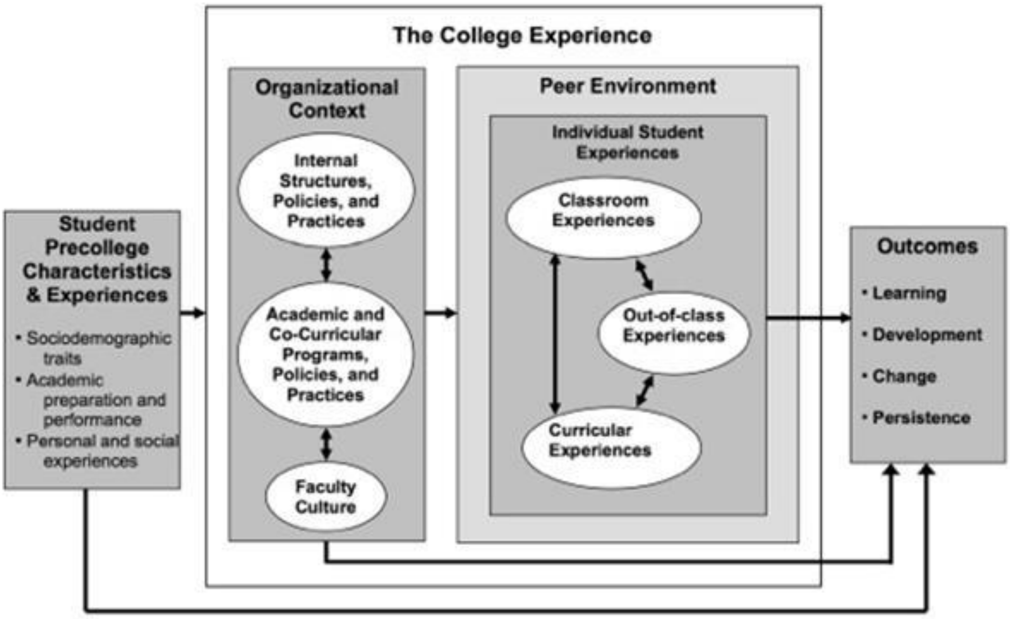

Figure 1-1. A Comprehensive Model of Influences on Student Learning and Persistence (Terenzini & Reason, 2013, p. 6)

Student experiences are also influenced by the organisational context of the institution the student attends (Terenzini, Ro, & Yin, 2010). The organisational context in this model is more specific and more directly related to the student experiences than abstract structural demographic features, such as size, kind of institution or admission selectivity, which appear to have little explanatory power for student learning (Pascarella & Terenzini, 2005; Terenzini et al., 2010).

The first set of factors in the organisational context that Terenzini and Reason (2013) identify, are the internal structures, policies and practices. It refers to factors in the organisation, such as the availability for educational development activities for teachers or the policies to support collaborations between different partners in education, like teachers, student affairs members and management. The second set of factors in the organisational context discerned by Terenzini and Reason (2013) is faculty culture. They assert that this culture is a subtle but significant factor in shaping student experiences. Dominant conceptions of what good education entails and ideas about the role of the teacher and the students therein are part of the faculty culture. The third set of factors –and the most significant for this doctoral dissertation- are the academic and co-curricular programmes, policies and practices. The curricular factors relevant in the organisational context pertain to the "formal" curriculum (Terenzini & Reason, 2005). The formal curriculum is what is officially intended, and the "enacted" curriculum, what is actually taught to the students (see Marsh & Willis, 1999). The curricular factors relevant for the individual student experiences pertain to the "received" curriculum, what students experience or how they perceive what is happening during the classes (see Kelly, 2009). The co-curricular factors are factors related to but not part of the curriculum. The three sets of factors are dependent on each other. The design of the curriculum is influenced by the internal structures, policies and practices and faculty culture, which is also hypothesised in the model of Stark and Lattuca (1997). Terenzini and Reason (2005) warn about the fact that the assumed influences described in the model are in reality not that linear. In addition they point out that, although the schematic representation suggests the possibility of a direct influence of the organisational context on student learning, the effect might be indirect and mediated by student experiences (Reason, Terenzini, & Domingo, 2006).

Concerning the organisational context it is important to note that it includes what Stark, Briggs, and Rowland-Poplawski (2002) refer to as "context", such as institutional and departmental characteristics and cultures. In addition, it also entails the curriculum and co-curricular programmes, policies and practices. This doctoral dissertation focuses on two elements of the conceptual framework presented above and their mutual relationship. The first is the organisational context and more precisely the curricular practices, as a factor of the academic and co-curricular programmes, policies and practices. The analysis of the curricular practices focuses on research integration practices. The second element is the learning outcomes, i.e. CT.

1.5 Context: Universities and University Colleges

The studies within this doctoral dissertation are conducted at different institutes of higher education in Flanders, the Dutch speaking part of Belgium. In Flanders higher education institutes are either university colleges or universities (Verhoeven, 2008).

The distinction between university colleges and universities is part of the discussion about research integration. Institutes of higher education differ in the position occupied by research. Finding common and non-confusing denominators for types of institutions of higher education is difficult (Kyvik, 2009; Kyvik & Lepori, 2010). In this doctoral dissertation "university colleges" are used for the higher education settings where doing research is relatively new and "universities" for higher education settings with a long research tradition, while acknowledging the oversimplification of the complexities involved (Bleiklie, 2003).

Historically the university colleges focussed on the preparation for a particular profession (Griffioen, 2013). Employability of graduates and partnerships with employers were high on the agenda (Prokou, 2008). The intention was to provide specialised occupational skills, rather than universal knowledge (Becher & Kogan, 1992). Recently the importance of a scientific knowledge base and the extended use of abstract vocabulary in the education of students has grown (Heggen et al., 2010). There seems to rise a common belief among academics and societal stakeholders that all teaching at the level of higher education should take place in a research and scholarly atmosphere (OECD, 1998; Heggen et al., 2010). It has become more commonly accepted that also in university colleges teachers' engagement in research and research integration is important for education (Heggen et al., 2010, Kyvik & Lepori, 2010). As a consequence teachers in university colleges are increasingly expected to do research and to integrate research into their teaching (Kyvik, 2009; Kyvik & Lepori, 2010; Witte, van der Wende, & Huisman 2008). The process, referred to as academic drift (Burgess, 1973) or academisation (Kyvik, 2009), has been contested (Heggen et al., 2010, Kyvik, 2009). There is the fear that a too strong focus on research might be detrimental to the professional preparation of students and the employability of the graduates (Kyvik, 2009). The place of research within curricula at universities might be less contested and considered as more "natural", in comparison to the university colleges (Lepori & Kyvik, 2010; Witte et al., 2008).

The described evolution of university colleges also counts for the university colleges in Flanders. Historically their core business was to train professionals but since the nineties the climate has gradually changed, with an acceleration since 2003 (Verhoeven, 2008; 2010). In Flanders education in the university colleges consists of two tracks. The professional track offers programmes leading to a professional bachelor's degree. The other track, the academic track, offers programmes leading to an academic bachelor's or master's degree. Universities, where research is from the outset part of its tasks, offer academic bachelor's programmes and (academic) master's programmes too. The implementation of the Bologna process created momentum for a harmonisation of the academic track at the university colleges with those of the universities (Verhoeven, 2008). This harmonisation is referred to as "the academisation process". Teachers should be more involved in research and all academic programmes should be based on research. To support this harmonisation, university colleges had to associate themselves with a university and the research funding for university colleges was raised: Between 2001 and 2007 the research budget of the university colleges was more than doubled although it is still only a fraction on the research budget of the universities (Verhoeven, 2010). In addition, special funding for the academisation was provided, in order to support the academic programmes of the university colleges to meet the same standards as the universities by 2013 (Verhoeven, 2010). Since 2013 the standards for quality assurance are equal for all academic programmes (VLUHR, 2013). And from the academic year 2013-2014 onwards the academic programmes of the university colleges become integrated into the university (Integratie Decreet, 2012). Given this evolution in Flemish higher education one can expect that research integration is a salient topic, certainly for the academic programmes of the university colleges.

The situation and the rationale for research integration might differ between universities and university colleges (Brown & McCartney, 1998), but the impact of the type of institution on research integration is understudied (Jenkins, 2004). While most research on research integration is performed in university settings (Heggen et al., 2010), this doctoral dissertation will mainly focus on university colleges. This exploration will illuminate research integration at university colleges.

1.6 Outline of this Doctoral Dissertation

The work is based on three data collections. The first data set consists of 45 professional bachelor's, academic bachelor's and master's programmes of different university colleges and one university. The second data collection took place at a university among freshmen (first year students) who took two tests on CT at the beginning of the academic year and took both tests again at the end of the academic year. The third data collection focuses on the first year of university colleges. Three academic and one professional bachelor's programme participated. Because research integration might differ between disciplines (Coate, Barnett, & Williams, 2001; Colbeck, 1998; Robertson & Bond, 2001; Rowland, 1996), all programmes participating in the study belonged to the same disciplinary area: hard sciences (Biglan, 1973). The teachers of the first year modules in these programmes participated in an interview. In addition, a developed CT test was administered twice to the freshmen of these programmes, once at the beginning and once at the end of the academic year. Only the data of three programmes could be used due to practical problems during the second administration of the fourth programme.

The doctoral dissertation tries to formulate an answer to the two research questions in eight chapters. The work is outlined in table 1-1.

Table 1-1

Overview and Structure of this Doctoral Dissertation

Chapter	Focus	Data
Chapter 1: Introduction	Research integration and CT	
Chapter 2: The selection of RR-learning outcomes and RI-approaches	Research integration	
Chapter 3: RR-learning outcomes: Programme patterns	Research integration	DC1
Chapter 4: RI-approaches: Prevalence and differences between modules	Research integration	DC1
Chapter 5: The relationship RR-learning outcomes and RI-approaches in modules	Research integration	DC3
Chapter 6: Measuring CT	CT	DC2
Chapter 7: Relationship between research integration and the development of CT	Research integration and CT	DC3
Chapter 8: Discussion	Research integration and CT	

Note. RR-learning outcomes = research-related learning outcomes, RI-approaches = research integration approaches, CT = critical thinking

The four chapters following the introduction concern research integration. Research integration refers to the way research is integrated into the learning environment and the way students are brought in contact with research. Since research integration is a multifaceted concept and since there is no agreement in the literature on what it exactly implies or includes, part I starts with an explanation of the selection of the two main variables (**Chapter 2).** In this chapter the selection of two different facets of research integration is explained and accounted for. First, research integration can be conceived as learning outcomes in relation to research. When students are for example expected to learn about research methods, it is the intended learning outcome which makes that teaching is considered as a manifestation of research integration. In this doctoral dissertation this will be labelled as "research-related learning outcomes" (RR-learning outcomes). Second, research integration can be understood as teaching approaches that make use of research. For example when students engage in research discussions. In this doctoral dissertation, teaching approaches which bring students in contact with research, are referred to as research integration approaches (RI-approaches). In the literature, the two facets are not always clearly distinguished. They are sometimes being used interchangeably or specific relationships between both are presumed. In this doctoral dissertation they are studied as two distinct facets.

In **Chapter 3** (Data collection 1) the prevalence of the RR-learning outcomes is studied. For this study the module descriptions of 45 programmes of different institutions of higher education were analysed to find indications for the pursuit of six RR-learning outcomes. The study investigates the prevalence of RR-learning outcomes in the programmes. In addition, it looks if different patterns in the prevalence of these RR-learning outcomes can be identified and whether these differences might be related to programme characteristics. In this chapter the formal curriculum is studied as a factor in the "Academic and co-curricular programmes, policies and practices" of the organisational context in the model of Terenzini and Reason (2013).

Chapter 4 concentrates on RI-approaches in the first year of the undergraduate curriculum in university colleges. The analysis is based on the interviews conducted in data collection 3. Teachers of 46 first year modules of four programmes within the same disciplinary field were interviewed about their way of teaching and their expectations towards students. The chapter aims at revealing the prevalence of RI-approaches in first-year programmes and studying whether different module types with specific combinations of RI-approaches could be discerned. Also in this study an element of the "Academic and co-curricular programmes, policies and practices" of Terenzini & Reason (2013) is studied, but whereas Study 1 focuses on the formal curriculum, here the focus is on the enacted curriculum as reported by the teachers of the modules.

In **Chapter 5** the relationship between RR-learning outcomes and RI-approaches is investigated. It also builds on the interview data collected in the third data collection. The chapter studies the relation between the different module types (identified in **Chapter 4**) and the research-related learning outcomes. In the existing categorisations of research integration a specific relationship between an approach and a learning outcome is often presumed. This chapter examines whether this assumption is reflected in practice.

Chapter 6 (Data collection 2) deals with CT. Since no validated instrument to assess CT in Dutch was available, two tests were translated and validated. The study investigates the psychometric quality of a translation of the Cornell Critical Thinking Test (CCTT; Ennis, Millman, & Tomko, 2005) and the Halpern Critical Thinking Assessment (HCTA; Halpern, 2007; 2012) in a sample of Dutch-speaking freshmen majoring in educational sciences. Based on the results of this study, a selection of items of both tests were combined to the Scipio. The Scipio test was used in Chapter 7.

In **Chapter 7** research integration and CT are combined. The study investigates the relationship between research integration and the development of CT. It uses the data of three programmes of the third data collection. In relation to the model of Terenzini and Reason (2013) the study investigates the effect of curricular features, as aspect of the organisational context, on the learning outcome CT.

Chapter 8 provides a general discussion about the results of the different studies. In addition, methodological issues and implications of the choices made in this doctoral dissertation are discussed and directions for further research are proposed. The chapter ends with theoretical and practical conclusions.

Chapter 2: Selection of Research Integration Approaches and Research-related Learning Outcomes[4]

Chapter 2

2.1 Introduction

Research integration into teaching, in short research integration, is one aspect of the research-teaching nexus. It concerns how research is integrated in the learning environment and how students are brought in contact with research.

At least two different facets of research integration can be discerned (Elen & Verburgh, 2008). First, research integration can be conceived as learning outcomes in relation to research. For example students are expected to learn about research methods or students have to learn disciplinary content. In this interpretation it is the intended learning outcome that makes that the learning environment is considered as a manifestation of research integration. In this doctoral dissertation this will be labelled as "research-related learning outcomes" (RR-learning outcomes). Different RR-learning outcomes will be italicised.

[4] The chapter is based on two conference papers and two reports.

Elen, J., Schouteden, W., Verburgh, A., & François, S. (2009). *Integratie van onderzoek in onderwijs: Ambities en realisatie binnen de associatie KU Leuven (IOO-project). Inhoudelijk eindverslag van OOF-project 2006/03 [Integration of research into teaching: Ambitions and realisation in the Association K.U.Leuven. Content report]*. Leuven, Belgium: Associatie KU Leuven.

Elen, J., Schouteden, W., Verburgh, A., & François, S. (2011). *Integratie van onderzoek in onderwijs: Realisaties en percepties (IOO² project) [Integration of research into teaching: Realisations and perceptions]*. Leuven, Belgium: Associatie KU Leuven.

Schouteden, W., Verburgh, A., & Elen, J. (2012, September). Schouteden, W., Verburgh, A., & Elen, J. (2012, Septemberb). *Lecturers in professional fields in higher education: Conception of research and it perceived importance of research skills for students' careers.* Paper presented at the annual conference of ECER, Cadiz, Spain.

Verburgh, A., Schouteden, W., & Elen, J. (2009, August). *A look at the prevalence of research-related educational goals in educational programmes of higher education*. Paper presented at the bi-annual meeting of EARLI, Amsterdam, The Netherlands.

Second, research integration can be understood as teaching approaches that make use of research. There are different types of learning environments in which students are confronted with research or research-like activities. For example students engage in research discussions or students help the teacher to collect data. In this doctoral dissertation, teaching approaches that bring students in contact with research, are referred to as research integration approaches (RI-approaches). Different RI-approaches will be underlined.

The literature does not always clearly distinguish these two facets . Sometimes they are used interchangeably and sometimes specific relations between both are presumed. In this doctoral dissertation the distinction is maintained in order to be able to study each separately. In this chapter the selected RR-learning outcomes and RI-approaches are explained and their relation to the literature is discussed.

Some parts of this chapter overlap with parts of the other chapters, as the chapters are (submitted) journal articles that can be read independently. Because space in journal articles is often too limited to extensively discuss relations between specifically elaborated selections and the literature, the opportunity was seized to discuss the elaborated selections for this doctoral dissertation and their background into more detail in this chapter.

2.2 RR-learning Outcomes

2.2.1 Seven RR-learning Outcomes

Learning outcomes refer to what teachers want their students to learn, they could also be called objectives or aims (Anderson & Krathwohl, 2001). RR-learning outcomes are learning outcomes concerning learning content, skills or attitudes related to research. Seven RR-learning outcomes were selected and listed in Table 2-1.

Table 2-1

Seven RR-Learning Outcomes

RR-learning outcome	Description *Example*
Results	Students have acquired knowledge from results of research. *The student acquires knowledge of scientific research on glacier systems.*
Underpinnings	Students have gained insight into methodological and theoretical underpinnings of research. *The student understands the importance of the used research method for the validity of the results.*
Practical research skills	Students have developed particular practical research skills. *The student can look up relevant scientific information about a given topic.*
Competence to be a researcher	Students have the competence to be a researcher. *The student can, starting from a problem situation, make and present an analysis about a particular economic-geographic entity, through the integration of written documents, numerical data, field work and statistical analyses.*
Critical thinking	Students have a critical attitude towards information, knowledge and knowledge construction. *The student can critically interpret historical data and apply this critical sense in the interpretation of the present.*
Curiosity	Students have a curiosity towards evolutions in the discipline. *The student is intrigued by the subject of the module and are stimulated to follow up the developments in the field.*
Practice	Student have the ability to use research results in new situations or to see the practical implications of research results. *In order to solve a problem situation, the student is able to look for relevant research results and apply them.*

2.2.2 **Description of RR-learning Outcomes**

The selection of RR-learning outcomes started from an analysis of the literature on research integration. For the identification and description, especially the following literature was used: Durning and Jenkins (2005), Griffiths (2004),

Healey (2005a), Healey (2005b; Healey & Jenkins, 2009), Neumann (1992), van der Rijst (2009), Visser-Wijnveen, van Driel, van der Rijst, Verloop, and Visser (2010), Visser-Wijnveen et al. (2012), Zamorski (2002), Zimbardi and Myatt (2012) (see Table 2-1). In Table 2-2 the aspects related to RR-learning outcomes are described, later on (in Table 2-6) the RI-approaches will be compared. Given that the distinction between RR-learning outcomes and RI-approaches is not always made in the discussed publications, some choices in the tables are somewhat artificial.
Table 2-2 shows that all selected RR-learning outcomes are present in the literature but no single publication mentions them all.

Results refers to the acquisition of knowledge of results of research. These research results can be results of research conducted by the teacher but also –and probably predominantly- research conducted by others.

Results is mentioned in almost all studied publications, although some restrict it to more advanced or recent knowledge (Neumann, 1992; Zamorski, 2002). It is part of the tangible nexus of Neumann (1992). Visser-Wijnveen et al. (2010; 2012) refer to this learning outcome as "Academic knowledge".

Underpinnings and *Practical research skills* concern respectively gaining insight into methodological and theoretical underpinnings of research and the development of practical research skills. These are skills relevant for doing research.

Both learning outcomes are related to "Research-oriented teaching" of Healey (2005b), in which teaching is focused to understand the research process as well as learning inquiry skills. Both are also related to "Developing student understanding of the research process" of Durning and Jenkins (2005). However, similar to Visser-Wijnveen et al. (2010; 2012), Zimbardi and Myatt (2012) and Zamorski (2002), a distinction between these two RR-learning outcomes is made. Gaining insight in the foundations of research (*Underpinnings*) does not automatically imply learning practical skills for doing research (*Practical research skills*). A teacher can aim at developing an understanding of the theoretical underpinnings of the method or the impact of it on the results without aiming at learning to apply a research method and interpret the results gained with the method.

Underpinnings relates to insight in underpinnings of research and not learning about methodology per se. Visser-Wijnveen et al. (2010; 2012) refer to it as "Divulge research" that implies making students aware of research and the processes it involves. It is therefore also connected with Zamorski's (2002) "Understanding more fully the complex and provisional relationship with research and knowledge".

Practical research skills aligns with Zamorki's (2002) "Learning about research methods and skills", Visser-Wijnveen's "Train researcher" and the development of research skills within Healey's (2005b) "Research-oriented" quadrant.

In the development of the selection of RR-learning outcomes a distinction was made between *Practical research skills* and the *Competence to be a researcher,* with the latter being more encompassing than the former. The RR-learning outcome *Practical research skills* merely confines to the development of skills to do research, whereas the *Competence to be a researcher* involves creativity and ingenuity. It implies an integrated set of skills, attitudes and knowledge. It refers to the competence to set up research and conduct it, with attention for developing a research question, developing the design, etc. It is similar to the "Academic disposition" of Visser-Wijnveen et al. (2010; 2012).

Moreover, in the literature the possibility to be involved in research or to conduct a research study is often included as a way to integrate research and to develop research skills or competencies. Durning and Jenkins (2005) for example mention that the "Developing students abilities to do research" can be done through courses on research methods or through research projects. But it is unclear what precisely is aimed at with those "Research projects". This also count for "Research-based teaching" that encourages students to undertake research (Healey, 2005b). It might be that the involvement in those research projects is intended to acquire a specific research skills as well as to develop the C*ompetence to be a researcher* but it might also be intended to improve students understanding of the underpinnings of research (*Underpinnings*). Because the intended learning outcome is not specified, it is not included in Table 2-2.

Critical thinking[5] refers to the development of a critical attitude towards information, knowledge and knowledge construction. It connects with Zamorski's (2002) idea about understanding more fully the complex and provisional relationship between research and knowledge, but it is broader as information is included as well. It focuses on critically examining the value of information or knowledge. *Critical thinking* is also an aspect of the "Academic disposition" of Visser-Wijnveen et al. (2010; 2012) or of the "Scientific research disposition" of van der Rijst (2009). It was decided to separate both because their relation is not reciprocal. When one has developed the competence to be a researcher, one is able to think critically. But the other way around, being able to think critically, does not imply mastering the competence to be a researcher. Similarly if a teacher's aims at developing *Critical thinking*, it does not imply that he wants to stimulate the *Competence to be a researcher*.

Curiosity refers to the development of curiosity towards evolutions in the discipline. Essential is the aim to create enthusiasm to monitor new developments and to feel the excitement or surprise to find something new or unexpected or to solve a problem. "Developing an appreciation of research" (Healey, 2005a) can be linked with it. It is related to the "Inclination to know", as one of the six aspects of a

[5] In this doctoral dissertation 'Critical thinking' is used in two distinction ways. It is used as a RR-learning outcomes, which is an intended learning outcome for the teacher. In addition, the development of critical thinking is discussed and studied. In order to minimise confusion, the RR-learning 'Critical thinking' is always written at length and in italics (*Critical thinking*); in all other instances 'Critical thinking' is abbreviated to 'CT'.

“Scientific research disposition” of van der Rijst (2009), that is the inclination to wanting to know more, to read about new findings and new ideas.

Practice refers to the ability of students to look for new research results to solve a problem situation or, the other way around, to see the practical implications or usefulness of a research result in their (future) professional practice. It is hence more than mere application of the content taught, students have to be able to add something new or to integrate new information. It is what is aimed at in the “Industry project model” of Zimbardi and Myatt (2012), namely “Solving complex practical problems”.

Table 2-2

The Selection of RR-learning Outcomes in Relation to the Literature

RR-learning outcome	During and Jenkins (2005)	Healey (2005a)	Healey (2005b; Healey & Jenkins, 2009)[1]	Neumann (1992)
Results	Bringing the content of research into the curriculum [when it refer to learning that content]		Research-led teaching: Learning about current research in the discipline	Tangible nexus [when it refers to the acquisition of knowledge and the recent facts]
Underpinnings	Developing student understanding of the research process		Research-oriented teaching: Developing research and inquiry skills and techniques [when it refers to understanding the research process]	Intangible nexus: development of research attitudes
Practical research skills	- Developing student understanding of the research process [when it refers to developing research skills] - Developing students abilities to do research	Developing research skills	Research-oriented teaching: Developing research and inquiry skills and techniques [when it refer to developing research skills]	Intangible nexus: development of research skills and attitudes
Competence to be a researcher	Developing students abilities to do research			Intangible nexus: development of research skills and attitudes
Critical thinking				Intangible nexus: to question and be critical about knowledge
Curiosity		Developing appreciation of research		
Practice				

Table 2-2 Continued

RR-learning outcome	Van der Rijst (2009)	Visser-Wijnveen , 2010-2012	Zamorski (2002)	Zimbardi and Myatt (2012)[1]
Results		Academic Knowledge	Gaining knowledge from recent research	Knowledge of the discipline
Underpinnings		Divulge research	Understanding more fully the complex and provisional relationships between research and knowledge.	Understanding how knowledge is produced in the discipline
Practical research skills		Train researcher	The gradual development of various research skills Learning about research methods and skills	- Understanding of methodological approaches - Develop research skills
Competence to be a researcher	Scientific research disposition	Academic Disposition		
Critical thinking	Scientific research disposition: To be critical	Academic Disposition (critical thinking)	Understanding more fully the complex and provisional relationships between research and knowledge.	
Curiosity	Scientific research disposition: To know			
Practice				Solve complex problems

Note. 1. The model van Healey (2005b; Healey & Jenkins, 2009) builds on Griffiths (2004). Because of the high similarity between both models, the one of Griffiths is not included in the comparison. 2. Zimbardi and Myatt (2012) identified five types of undergraduate research experiences. They did not explicitly make a list of learning outcomes, but the learning outcomes they mention are placed in the table.

2.2.3 **Description of the Specifications of Practical Research Skills**

Practical research skills is a broad learning outcome, implying different interpretations. When talking about aspects of research, teachers mention different aspects which can be considered as a different practical skill to do research (Schouteden, Verburgh, & Elen, 2013). They can be translated into learning outcomes, as specifications of *Practical research skills.* They are listed in Table 2-3, together with a description and an example.

The first six research skills are closely related to traditional research steps (Cohen, Manion, & Morrison, 2007; Mertens, 2010; Nisbet & Entwistle, 1970). They are comparable to the facets of research identified in the research skill development framework of Willison and O'Regan (2007). Their framework is intended to guide research skills development of students. *Formulating a research question* is similar to their first facet "Embark and clarify" as it implies generate research questions and hypotheses. Their second facet "Find and generate" refers to collecting required information and data. This is partly what is meant with *Finding literature* and partly with *Collecting data*. The difference between both is that in the former existing information is meant, as an orientation on the topic while in the latter new data is collected. Although the skill of *Developing a design* is absent in the framework of Willison and O'Regan (2007), it is often stressed or split up in several steps in overviews of research steps (Cohen et al., 2007; Mertens, 2010; Nisbet & Entwistle, 1970). *Analysing data* and *Formulating a conclusion* is similar to "Organise and manage" and "Analyse and synthesise" although in our classification the analyse-aspect of "Analyse and synthesise" is part of *Analysing data*. The "Communicate and apply" of Willison and O'Regan (2007) is a combination of *Formulating a conclusion* and *Report*. The distinction between both specifications was made here because *Formulating a conclusion* demands an abstraction of information gathered. A

Table 2-3

Specifications of Practical Research Skills (Based on Schouteden et al., 2013)

Specification	Description *Example*
Formulating a research question	Students have developed the ability to formulate research questions and hypotheses. *The student can formulate a question that can be investigated.*
Finding literature	Students are able to retrieve the relevant literature. *The student can find scientific literature in a database given a question and valid keywords.*
Developing a design	Students are able to develop a suitable design for a given question. *The student can conceptualise a study.*
Collecting data	Students are able to collect data according to the scientific standards. *The student can observe the behaviour of a psychiatric patient in a systematic way.*
Analysing data	Students are able to conduct a data analysis. *The student is able to do a valid statistic analysis of the given data.*
Formulating a conclusion	Students are able to draw a conclusion. *The student can formulate a conclusion based on the question and the data given.*
Report	Students are able to communicate their activities or conclusions. *The student can write an account of what she has done during the lab session.*
Research attitude	Students have developed a research attitude. *The student does not accept statements as given but looks for supporting evidence.*
Systematic way of reasoning	Students are able to think logically and apply a heuristic way of reasoning appropriate in the discipline. *The student can approach a problem in a systematic way in order to come to an answer.*

conclusion transcends the given. The core of *Report* is the act of communication and it can be confined to communicate results without interpretation.

A *Research attitude* is considered as a *Practical research skill* and not as a direct indicator of the *Competence to be a researcher* because having the attitude to question things or to look further than the given information is not the same as being able to investigate the questions posed.

A *Systematic way of reasoning* is a skill that permeates in all research steps and therefore identified as a separate specification. It refers to learning outcomes such as "Perceive relations and patterns", "Apply theory

and balance different perspectives" which are learning outcomes ascribed to undergraduate research (Lopatto, 2009).

2.2.4 **Description of the Specification of Critical Thinking**

CT is a multifaceted concept (Halpern, 1998) and the RR-learning outcome *Critical thinking* can also imply different specifications. Teachers talking about CT, refer to different aspects of it (Schouteden et al., 2013). More specifically, four different RR-learning outcomes about critical thinking can be identified (Table 2-4).

Critical thinking towards oneself refers to be able to question the personal way of acting and the personal frame of reference. It is comparable to the "Self-critical attitude" of van der Rijst (2009), implying being critical towards one's own ideas.
Critical thinking towards information implies being able to critically assess information and the way it was developed. It is related to the "Critical attitude towards others" and "Critical attitude towards observations and experiments" of van der Rijst (2009).

Table 2-4

Specifications of Critical Thinking (Based on Schouteden et al., 2013)

Specification	Description *Example*
Critical thinking towards oneself	Students are able to think critically towards themselves, their own way of acting and the own frame of reference. *The student can critically analyse the own behaviour during an internship according to the theoretical framework used in the module.*
Critical thinking towards information	Students are able to think critically towards information and the development of information and knowledge. *The student can critically interpret historical data and apply this critical sense in the interpretation of the present.*
Conscious of perspective others	Students are conscious of the perspective or framework of others. *The student can put herself in the perspective of each actor in a given problem.*
Able to handle uncertainty	Students are able to stand and cope with uncertainty. *The student can formulate a conclusion based on mixed and incomplete information.*

C*onscious of the perspective others* refers to being able to see the perspective or frame of reference of others. *Able to handle uncertainty* includes the ability to stand conflicting information and accept that each position contains aspects of truth. These last two interpretations are closely related to metacognition (Efklides, 2008) and advanced epistemological beliefs (Perry, 1970). Reflective thinkers in the model of King and Kitchener (1994) fully master these skills, as they recognise uncertainty and the construction of knowledge while also acknowledging that knowledge must be understood in relation to the context in which it was generated.

2.3 RI-approaches

2.3.1 Selected RI-approaches

In this doctoral dissertation RI-approaches are understood as teaching approaches in which students come in contact with research or research-like activities. This conception of RI-approaches is broader than undergraduate research, in which students are involved in authentic research. It is generally accepted that research integration is more encompassing than undergraduate research (e.g., Brew, 2013; Elsen, Visser-Wijnveen, van der Rijst, & van Driel, 2009; Healey, 2005b; Levy & Petrulis, 2012; Spronken-Smith & Walker, 2010; Visser-Wijnveen et al., 2010; 2012). Nevertheless, there is discussion about what counts as research integration. Even what is considered as undergraduate research varies. Undergraduate research can be defined as "an inquiry or investigation conducted by an undergraduate student that makes an original intellectual or creative contribution to the discipline" (Council on Undergraduate Research (CUR.org)), while for others the original contribution is not essential (Brew, 2010). In this doctoral dissertation a broad interpretation of RI-approaches is opted for.

Based on the work of Elen et al. (2011) a classification with two dimensions was developed: research processes and research outcomes. While others (Elsen et al., 2009; Healey & Jenkins, 2009) used research results and the research processes as opposite positions on one dimension, here they are used as two independent dimensions because both can simultaneously be relevant. This contributes to distinguish between different RI-approaches and to avoid overlap. It makes the different RI-approaches more delineated. It also adds to the possibility to study differences and similarities in the effects of different approaches.

For the research processes, six alternatives are considered: 1. No focus on the research background, 2. Research background is mentioned, 3. Research background is explained, 4. Research methods are explained, 5. Segments of a research process are run through; and 6. A full research cycle is made.

The distinction between the first three possibilities is made based on their different face validity and because of the assumption of a differential effect on student learning. The appreciation of approaches without focus on the research background is different than approaches with a focus on the research background. Some authors might not consider the former as a genuine research experience (e.g., Beckman & Hensel, 2009).

The difference between the first four and the last two concerns whether the content is "given" or when students have to produce (some) content themselves, which is for some authors the demarcation line to talk about a research experience (e.g., Zimbardi & Myatt, 2012; Brew, 2013). The distinction between "Segments" and "Full study" is also made to better distinguish between different practices so that the practices within one approach are more similar.

The dimension research outcomes concerns the intended implications of the outcomes of the research activities of students. There are two possibilities: research outcomes are relevant for the student or functional for the discipline. The dimension relates to the tension of originality (Beckman & Hensel 2009; Elsen et al., 2009; Levy & Petrulis, 2012). This tension concerns the question to whom the outcomes of the conducted research activity should be original: the student or the discipline. For example, repeating a classic experiment leads to results that are new for the students but not at all to the discipline (or else they are original to the student but not original to the discipline). Here these activities are considered as "relevant for the student", they are not intended to add something to the discipline but the activities are oriented towards student learning. When a research activity is functional for the discipline, it implies that it aims at contributing to the knowledge base of the discipline. This functionality for the discipline, does not exclude relevance for student learning. The distinction refers to the main orientation.

The dimension research outcomes is similar but different from the distinction made by Visser et al. (2010) on the orientation of the research-teaching nexus towards teaching or towards research. In their distinction, when the nexus is oriented towards teaching, it concerns student learning that benefits from research. When the nexus is oriented towards research, it is the research of the academic that benefits from teaching, such as the teachers (as researcher) being stimulated to think thoroughly because of students questions . So for Visser et al. the benefit for research concerns mainly the academic as person and it is not as broad, as in our case, the discipline as a whole.

Combining the two dimensions, eight meaningful positions concerning research integration are specified (Table 2-5). The dimension of research outcomes becomes relevant only when students are involved in doing (segments of) research.

Table 2-5

RI-approaches by The Dimensions Research Processes and Research Outcomes

	Research outcomes	
Research processes	Relevant for students	Functional for discipline
No focus on research background	Facts	/
Research background is mentioned	Scientific facts	/
Research background is explained	Research-based facts	/
Research methods are explained	Research methods	/
Segments of research are run through	Segments-outcomes relevant	Segments-outcomes functional
A full research cycle is made	Full study-outcomes relevant	Full study-outcomes functional

The RI-approaches solely refer to the way students are brought in contact with research. Within one RI-approach there are different instructional techniques and different expected learning activities of students possible. Moreover, there is no assumed relation with what students have to learn. This implies that with one RI-approach different RR-learning outcomes can be aimed at, and one specific RR-learning outcome can be aimed at using different RI-approaches. Although the dimension research outcomes refers to student learning, it does not imply a specific learning outcome. It only states that the activity is primarily oriented towards student learning or to the advancement of the discipline, it does not specify the intended learning outcome.

2.3.2 Description of the RI-approaches

Facts

Students are confronted with descriptions of research results as if they were facts. The research background of the results is not mentioned. This approach can be experienced in various settings, for example, attending a traditional lecture, listening to an explanation of a theory by means of examples, participating in a Socratic conversation with a teacher posing questions and students answering, watching a documentary film, reading a text or explaining a given text to peers. The absence of any mentioning of the research background of the findings is the key feature.

Scientific facts

Students are confronted with research results. Here the research background is touched but no further explanation of the research is given. This approach can be experienced in similar settings as Facts, but with touching the research background. For example, during the lecture or the documentary film the speaker says that research has revealed the findings, drops the name of a researcher or the scientist, or a text includes in-text references to ground the statements made. Distinctive is the mentioning of the research background without further explanation.

Research-based facts

Students are confronted with research results and the research behind those results. The scientific background of the facts is explained. Explaining the results of a study and why this is relevant for the theory under study is an example of this approach. Other examples are a teacher showing a documentary film exemplifying the experiment underlying the theory being studied or a teacher explaining her students why the value of a particular parameter changes year after year because of evolutions in the discipline. In this approach students are confronted with the results of research and their background.

Research methods

Students are confronted with methods of conducting research. For example, they have to read an explanation about an analysis technique and how it should be used. Another possibility is that a teacher demonstrates how to handle a machine to do measurements. The focus is on explanations of research methods.

Segments- outcomes relevant

Students are actively involved in doing one or more research steps. The research activity is not oriented towards advancing the discipline but towards student learning. This approach can be experienced in diverse settings, e.g., during lab-sessions students have to repeat a classic experiment, students receive a question and they have to think about a method to investigate the question, students have to look up information about a given topic, or students have to select and apply an appropriate analysis technique to answer a given question. Students are actively involved in doing segments of research, and the activity and outcomes are relevant for student learning.

Segments- outcomes functional

Students are involved in doing one or more research steps, as in Segments- outcomes relevant. However, in this category results are not only relevant for the students but also intended to add to the advancement of the discipline. For example, students have to fine-tune a distillation machine, so that the teacher can use it in her experiments, students have to administer questionnaires of a research study of the teacher, or students have to perform statistical analyses on real research data. Students are actively involved in doing segments of research, and their activity is intended to be functional for the discipline.

Full study- outcomes relevant

Students are involved in a complete research cycle and the results of the study are relevant for the students. For example students have to do a small literature review but the results of the review are not important to or are already known by the teacher. Students' research is oriented towards their learning.

Full study- outcomes functional

Students are also involved in a complete research cycle, but in contrast to the previous approach, the results are intended to advance the discipline. Here students become members of the research community. For example students have to do a literature review that helps the teacher to develop a hypothesis for future experiments. The research students are involved in, is functional for the advancement of the discipline.

2.3.3 Relation with the Literature

The distinction between Facts, Scientific facts, and Research-based facts is generally not made in the literature (Table 2-5). The descriptions are often not specific enough to know precisely what authors mean. For example, "Research- led" (Healey, 2005b; Healey & Jenkins, 2009) refers to teaching based on the transition model and teaching that is structured around subject content. This could imply that the research background of the subject content is explained in detail, but it could also be the opposite in that only the content as such is mentioned without any research background. Some descriptions suggest that the research behind the facts is mentioned, for example in the approach of "Bringing data and findings from staff in the curriculum" (Healey, 2005a).

The distinction between <u>Research-based facts</u> and <u>Research methods</u> is based on the content at stake in the approach. In the former the content is the results of research, in the latter it is the methods of research, but in either RI-approach, the content "given". <u>Research methods</u> relates to the "Research methodology course" of Zamorski (2002) or the "Methodology modules" of Durning and Jenkins (2005). The "Talking about research conducted by the academic" of Durning and Jenkins (2005) could refer to <u>Research-based facts</u> or <u>Research methods</u> because it is conceivable that the content of the talk is about the results or about the methodology.

The distinction between research activities relevant for student learning and functional for the discipline is not always made. The "Engage in research activity" of Zamorksi (2002), "Research project" of Healey (2005a) and "Research project" of Durning and Jenkins (2005) as well as the research discussions of "Research-tutored teaching" or "Research-based teaching" of Healey (2005b) could all refer to both. For example, the research discussions in "Research tutored" (Healey, 2005b) could be restricted to student learning, but it could also be intended for the advancement of the discipline, for example when the academic wants to discuss his preliminary findings with students to refine her own thinking. It relates to <u>Segments</u> because students are involved in one research step only. A "Research project" could be intended to familiarise students with a specific research method but it could also be intended to collect data for a research study of a doctoral student of the teacher.

The RI-approach <u>Segments- outcomes relevant</u> shows similarities with Zimbardi and Myatt's (2012) "Methods course model", because their students are actively involved in one or more research steps and the activity is aimed towards student learning. It also relates to "Inquiry learning" of Visser-Wijnveen et al. (2010; 2012) because, in their conception of inquiry learning, students are involved in segments of research without doing actual research.

Chapter 2

The RI-approach Full study- outcomes relevant is similar to the “Inquiry project model” of Zimbardi and Myatt (2012), and the “Simulation” of Visser et al. (2010; 2012). These authors refer also to activities that are comparable to a full research cycle, but not with the intention to add something to the discipline but to stimulate student learning.

The RI-approach Full study- outcomes functional is connected with the “Apprenticeship model” of Zimbardi and Myatt (2012), “Undergraduate research” (as defined by the council of undergraduate research (Cur.org)) and “Participation” of Visser-Wijnveen et al. (2010; 2012). However, in these classifications, it is often unclear if teaching practices where students are only involved in a few steps of a research process, are considered as a genuine examples.

Table 2-6

The Selection of RI-approaches in Relation to The Literature

RI-approaches	Durning and Jenkins (2005)	Healey (2005a)	Healey (2005b; Healey & Jenkins, 2009)	Neumann (1992)
Facts	Bringing content of research into the curriculum		Research-led teaching: Teaching is constructed around subject content and teaching is based on the transmission model	Use of scientific research results
Scientific facts	Bringing content of research into the curriculum		Research-led teaching: Teaching is constructed around subject content and teaching is based on the transmission model	Use of scientific research results
Research-based facts	Bringing content of research into the curriculum Talking about research conducted by the academic	Bringing data and findings from staff in the curriculum	Research-led teaching: Teaching is constructed around subject content and teaching is based on the transmission model	Use of scientific research results
Research methods	Talking about research conducted by the academic Methodology modules			
Segments- outcomes relevant	Research project	Research projects	- Research-tutored teaching: Teaching to stimulate student discussing and writing about research results - Research-based teaching: Teaching is constructed around inquiry-based activities	
Segments- outcomes functional	Research project	Research projects	- Research-tutored teaching: Teaching to stimulate student discussing and writing about research results - Research-based teaching: Teaching is constructed around inquiry-based activities	
Full study- outcomes relevant	Research project	Research projects	Research-based teaching: Teaching is constructed around inquiry-based activities	
Full study- outcomes functional	Research project	Research projects	Research-based teaching: Teaching is constructed around inquiry-based activities	

Table 2-6 (Continued)

RI-approaches	Undergraduate research (www.cur.org)	Visser-Wijnveen (2010; 2012)	Zamorski (2002)	Zimbari and Myatt (2012)
<u>Facts</u>		Learning about research: lecturing and literature reading	Taught courses	
<u>Scientific facts</u>		Learning about research: lecturing and literature reading	Taught courses	
<u>Research-based facts</u>		Learning about research: lecturing and literature reading	Taught courses	
<u>Research methods</u>			Research methodology course	
<u>Segments- outcomes relevant</u>		- Inquiry learning: containing analysis, assignments, discussions, reporting. Students learn in a research-like way without doing actual research - Simulation	Engage in research activity	Methods course: Involving students in data analysis and/or critically analysing authentic research articles for the methodological approach and the appropriateness of the conclusions drawn.
<u>Segments- outcomes functional</u>	Undergraduate research	Participation	Engage in research activity	Apprenticeship model: Students "work under the direct supervision of a researcher on a question directly related to the current research interests of the researcher" (p. 7) (Industry project: Students solve authentic industry-related problems)*
<u>Full study- outcomes relevant</u>		Simulation	Engage in research activity	Inquiry project: Students develop identify a problem, develop a research design and conduct experiments or do analyses on existing data resulting an a written report,
<u>Full study- outcomes functional</u>	Undergraduate research	Participation	Engage in research activity	Apprenticeship model: Students "work under the direct supervision of a researcher on a question directly related to the current research interests of the researcher" (p. 7) (Industry project: students solve authentic industry-related problems)*

Note. *Depending on the similarity between the industry project and research it could or could not be considered as RI-approach in our classification.

Chapter 3: RR-learning Outcomes: Programme Patterns[6]

Integration of research into teaching is considered a desirable characteristic of higher education, and valuable for student learning. Nevertheless, not much is known about the prevalence of research integration in students' everyday experiences within higher education. This article reports on an empirical study on the prevalence of six RR-learning outcomes in 45 educational programmes in Flanders. The study aimed at identifying patterns in the prevalence of these learning outcomes at programme level. For each programme, a profile was identified. A cluster analysis revealed three different patterns: a low-attention pattern (N = 21), a results pattern (N = 16) and a critical-thinking pattern (N = 8). The prevalence of these patterns differs between distinct types of programmes, but not between disciplines.

3.1 Introduction

The relationship between teaching and research touches the core of higher education (Clark, 1997). The cohabitation of teaching and research in the same institution is considered as enriching for both (Taylor, 2007). However, empirical evidence of a close link between research and teaching is limited. A seminal review of Hattie and Marsh (1996) on the relationship between research output and teaching

[6] This chapter was published within Teaching in Higher Education as Verburgh, A.L., Schouteden, W., & Elen, J. (2013c). Patterns in the prevalence of research-related goals in higher education programmes. *Teaching in Higher Education, 18, 298-310.*
To improve the readability of the doctoral dissertation 'goals' were replaced by 'learning outcomes' and 'vocational' was replace by 'professional'

satisfaction revealed a slightly positive, but close to zero, correlation. They conclude "that the common belief that research and teaching are inextricably entwined is an enduring myth." (p. 529). Since their review, much research has been undertaken, in an attempt to unravel this complex and multilayered relationship. Often these studies are oriented towards understanding the relationship, and revealing how research may support teaching (Prince, et al. 2007). There is a strong conviction amongst teachers, students and administrators about a positive relationship between teaching and research (e.g., Neumann, 1994; Rowland, 1996; Leisyte et al., 2009). Most commonly, research is considered to be important for the quality of teaching, while the idea of teaching being important for research is less frequently expressed (Taylor, 2007). In addition, there have been numerous projects on "good practices" concerning research integration to inspire teachers (e.g., Malcolm, 2008). Studies of students' research participation programmes (Craney et al., 2011; Hunter et al., 2007) show personal and professional- research-related- growth amongst the participating students.

These studies address to a certain extent the point of Hattie and Marsh (1996) that the level to which research actually becomes integrated within teaching is hardly investigated. The studies shed light on the perceived importance of research integration and possible ways to do such an integration. In addition, they demonstrate the beneficial effect of research participation. Nevertheless, it remains unclear how intensively students are confronted with research-integration throughout their curriculum. Or put differently, it is unknown what the research profile of teaching is. It may be that the constraints of daily practice drive teaching and research apart (Leisyte et al., 2009) and hence hinder teachers in integrating research into teaching to the extent they desire.

The question of how often students experience research integration throughout their curriculum is relevant because the relationship between teaching and research is assumed to be an important - or even a defining - feature of higher education, and the integration of research into teaching can be considered an important manifestation of that relationship (Taylor, 2007). Therefore, it would be valuable to know to what extent research integration is a part of students' daily educational experience, and to know how often the endeavours of teachers are translated into educational practice. A minimal indicator of this challenging question could be the extent to which teachers aim to achieve RR-learning outcomes. RR-learning outcomes are objectives concerning learning content, skills or attitudes related to research. This study aims to illuminate the prevalence of educational learning outcomes related to research in modules and programmes. It investigates whether different patterns in the prevalence of these RR-learning outcomes can be identified. In addition, the study aims to explore some expectations regarding these patterns, based on factors identified in the existing literature, namely type of programme and discipline.

In the setting for this study -Flanders, Belgium- programmes are defined as structural entities in the educational scope of an institution leading to a degree. Modules are smaller units within a programme. Within a programme, students have to follow several modules. It is pre-determined which modules students have to follow, and between which modules they may choose. Upon registration at an institution, students select a specific programme.

The factor *type of programme* pertains to two factors; the level of the programme, and the orientation of the programme. First, the *level of the programme*, empirical studies suggest that the relationship between teaching and research differs either between different levels of programmes (bachelor/undergraduate versus master/postgraduate versus PhD-students) or between first year students and not-first year students. Teachers perceive that teaching and research activities are more intertwined and the distinction between the two becomes blurred at postgraduate level (Neumann, 1992). Teachers consider research integration as more natural (and hence more common) in the more advanced classes of the bachelor's and master's programmes compared to the earlier years (Elen & Verburgh, 2008; Taylor, 2007). Second, the *orientation of a programme* might be of influence. Higher education programmes can be more professionally, or more academically, oriented. Programmes with a professional orientation are intended to prepare students for a particular occupation or profession which requires an extended study in higher education. Programmes with an academic orientation are more theory-oriented, or oriented towards a more general, abstract introduction to a scientific discipline (Huisman & Kaiser, 2001). The distinction between the orientation of programmes can follow the organisational structure of higher education in a country, but the distinction is often not clear-cut, for example in unitary systems (Huisman & Kaiser, 2001). Research integration is assumed to be more prominent in academic programmes than in professional ones.

Discipline is included as a variable in the study because several studies have concluded that it affects the relationship between teaching and research. In this respect, the distinction between hard and soft disciplines is often made (Biglan, 1973). Research has shown that teachers of soft sciences, such as humanities, find it easier to integrate research into their teaching than their colleagues from other disciplines, such as sciences or mathematics (e.g., Coate et al., 2001; Robertson & Bond, 2001; Rowland, 1996). Kyvik and Smeby (1994) hypothesised that these differences may be attributable to the higher agreement on paradigms in "hard" sciences in comparison to "soft" sciences, and to the knowledge structure of disciplines: in hierarchical disciplines (hard sciences) students need first to master an extensive knowledge base before they thoroughly understand the discipline and can make sense of research.

3.2 Method

3.2.1 Data Set

The sample of programmes is an opportunistic sample of 45 programmes of the KU Leuven Association, in Flanders, Belgium. The KU Leuven Association is a group of 12 university colleges and the University of Leuven, representing about 44% of the total number of higher education students in Flanders. About 13% of all (regular, Dutch) programmes of the association from the different institutions volunteered to participate in the study because research integration into teaching is an important topic for them.

Characteristics of the data set: type of programme , discipline and same degree programmes

Three types of programmes are included: professional bachelor's, academic bachelor's and master's programmes (see Table 3-1). In Flanders, programmes have either a professional or an academic orientation. Professional programmes are organised by university colleges, which are non-university institutions (Verhoeven, 2010). They are bachelor's programmes of 180 ECTS-credits (European Credit Transfer and accumulation System, where 60 credits equal one full year of study) and are scaled at 5B medium level in the International standard classification of education of the OECD (ISCED-97 classification system). The main focus of academic bachelor's programmes is to prepare students for a master's degree, which have, by law, an academic orientation. Academic programmes are organised by university colleges and universities. Academic bachelor's programmes are programmes of 180 ECTS-credits, master's programmes consist of either 60 or 120 ECTS-credits. They are scaled at 5A level in the ISCED-97 classification system. For example, within the domain of chemistry, the professional bachelor's in chemistry prepares students to become chemical laboratory assistants, whereas the master's in chemistry introduces the student to the research field of chemistry and prepares them to be a chemist, able to conduct research.

The programmes were identified as belonging to a hard or soft discipline following Biglan (1973) (see Table3-1). About two-thirds of the programmes in the sample belong to a hard discipline, and one-third to a soft discipline. This distribution is comparable across the different programme types within the sample. There is an underrepresentation of the soft sciences compared to the programmes organised by the KU Leuven Association.

Table 3-1

Number of Programmes by Discipline and Type of Programme (Percentage of the Overall Regular Educational Offer)

	Hard	Soft	Total
Professional bachelor	15 (26,0%)	7 (10,0%)	22 (17,1%)
Academic bachelor	6 (13,3%)	3 (7.7%)	9 (10,7%)
Master	10 (15,0%)	4 (6.3%)	14 (10,8%)
Total	31 (18,2%)	14 (8,1%)	45 (13.1%)

As the sample contains programmes from different institutions, it is possible to compare programmes of different institutions leading to the same degree. To guarantee anonymity only the broad disciplinary area is mentioned. In the dataset, ten programmes offered the same professional bachelor's degree in biomedical sciences, three offered the same professional bachelor's degree in sciences, and three offered the same academic bachelor's degree in applied sciences. The comparison of programmes offering the same degree makes it possible to see whether the prevalence of RR-learning outcomes is related to the nature of the degree. In that case, large differences between different programmes offering the same degree would not be expected.

3.2.2 RR-learning Outcomes

For the study, six RR-learning outcomes were selected. Learning outcomes refer to what teachers want their students to learn, they could also be called objectives, aims, guided outcomes (Anderson & Krathwohl, 2001). The selection of the learning outcomes is based on the literature on types of research integration (Durning & Jenkins, 2005; Griffiths, 2004; Healey, 2005b; Zamorski, 2002) and on interview studies on research-integration among teachers (Elen & Verburgh, 2008; Neumann, 1994). The pursuit of these learning outcomes is considered as an indication of the attention devoted to teaching students content, skills or attitudes related to (doing) research.

The six RR-learning outcomes are

- *Results*: Acquiring knowledge from results of research;
- *Underpinnings*: Gaining insight into methodological and theoretical underpinnings of research;
- *Practical research skills*: Developing particular practical research skills;
- *Competence to be a researcher*: Developing the competence to be a researcher;
- *Critical thinking*: Developing a critical attitude towards information, knowledge and knowledge construction; and
- *Curiosity*: Developing a curiosity towards evolutions in the discipline.

The *Competence to be a researcher* is more encompassing than *Practical research skills* where a single skill is aimed at, such as being able to report findings. *Critical thinking* focuses on critically examining the value of given information. *Curiosity* refers to the willingness of students to follow future developments in the discipline, to the attempt of teachers to intrigue students about what is unknown in the discipline and how it might evolve (See Verburgh et al., 2009 for a detailed description).

3.2.3 The Profile of Prevalence of RR-learning Outcomes

The identification of the profile of the prevalence of RR-learning outcomes of a programme consists of two distinct steps. In the first step, the prevalence of the six

RR-learning outcomes of each individual course in a programme was assessed by looking at the course descriptions. The course description is a short outline of what a course is about, and how it will be approached. Possible elements of a course description include the aims of the course, the content, the teaching approach, the evaluation or the number of credits. The profile of the prevalence of RR-learning outcomes is hence based on the formal curriculum, the programme set on paper (Van den Akker, 2003). This choice is based on the assumption that the public information informs about the enacted curriculum (what actually happens in the class) in a lucid way because of the specific official status of these course descriptions in Flemish higher education. Course descriptions are considered as the contract between the institution and the student (Education and examination regulations at the KU Leuven, 2011). And they are part of the documentation for external quality assurance initiatives.

In total, 1949 modules received a score for each of the six learning outcomes. The scoring is based on a strict coding protocol (Verburgh et al., 2009). Scores could be 0, 1 or 2. Score 0 was assigned if there was no sign of the pursuit of that learning outcomes in the course description. Score 1 was given if there was an indication of the pursuit of the learning outcome, but it could not be stated with certainty. Score 2 was given if there was a clear indication of the pursuit of the learning outcome. In Table 3-2 examples of learning outcomes receiving a score 2 are listed.

Table 3-2

Example of Learning Outcome Descriptions which are Rewarded with Score 2

Learning outcome	Example
Results	Students acquire knowledge of scientific research on glacier systems.
Underpinnings	Students understand the importance of the used research method for the validity of the results.
Practical research skills	The student can observe, analyse, interpret and report on the behaviour of a psychiatric patient.
Competence to be a researcher	The student can, starting from a problem situation, make and present a SWOT-analysis (Strengths, Weaknesses, Opportunities and Threats- analysis) about a particular economic-geographic entity, through the integration of written documents, numerical data, field work and statistical analyses.
Critical thinking	Students learn to critically interpret historical data and apply this critical sense in the interpretation of the present.
Curiosity	Being confronted with real environmental problems, students are intrigued by the issue and stimulated to follow up the developments in the field.

The reliability of the scoring of the six RR-learning outcomes is checked by a weighted Cohens' kappa for each of the learning outcomes. It is based on 213 randomly selected modules spread over 19 programmes, independently scored by two raters. It ranges between substantial to almost full agreement, depending on the learning outcome (Cohen, 1988). The weighted Cohens' kappa is the lowest for *Underpinnings*, but is still substantial (0.61). The other learning outcomes have an interrater reliability between 0.75 and 0.96 (*Results*: 0.76; *Practical research skills*: 0.87; *Competence to be a researcher:* 0.96; *Critical thinking*: 0.88; *Curiosity*: 0.83).

In the second step, the profile of the prevalence of RR-learning outcomes at the programme level was calculated based on the weighted scores of the individual modules in a programme. The weight depends on the number of ECTS-credits awarded for a fixed course, and on the number of credits students have to select in case of electives. This approach allows to account for the relative importance of a course in a programme.

3.2.4 Expectations to Explore

Based on the literature, expectations regarding the prevalence of the RR-learning outcomes differ according to type of programmes and discipline. A higher prevalence of RR-learning outcomes can be expected in the master's than in the bachelor's programmes. For the bachelor's programmes, academic bachelor's programmes are expected to concentrate more on RR-learning outcomes than professional bachelor's, and more attention to RR-learning outcomes can be expected in the soft sciences than the hard sciences, which may concentrate more on gaining knowledge of research results.

3.2.5 Analysis Techniques

To identify different patterns in the profiles of the programmes, a *K*-means cluster analysis was performed. Before the analysis, visual representation of all programme profiles were compared (Everitt, 1993). To determine the number of clusters, a hierarchical cluster analysis was performed. Where applicable, a one-way between-groups multivariate analysis of variance (MANOVA) was undertaken to investigate possible significant differences between the patterns for each of the learning outcomes. When an individual learning outcome did not meet the preliminary assumptions a Kruskal-Wallis test was performed instead. χ^2-analyses were undertaken to test for significant differences between the patterns in relation to the type of programme and to the discipline, because they are categorical variables (Howell, 1992).

3.3 Results

3.3.1 Three Patterns

Based on the distances in the dendogram of the hierarchical cluster analysis, a 3-cluster solution was adopted (see Table 3-3). Next, a MANOVA was performed. Preliminary tests showed that *Underpinnings* and *Curiosity* did not meet the assumptions of normality and univariate outliers. Therefore they were excluded from the MANOVA, and for these two learning outcomes a Kruskal-Wallis test was performed instead. There was a statistically significant difference between the three patterns concerning the four included learning outcomes, $F(8, 78) = 19.65$, $p < .001$; $\eta^2_p = .67$. When looking at the four learning outcomes separately, they all were significantly different from one other ($F_{Results}(2, 42) = 14.68$, $p < .001$; $\eta^2_p = .41$; $F_{Practical\ research\ skills}(2, 42) = 43.59$, $p < .001$; $\eta^2_p = .68$; $F_{\text{competence researcher}}(2, 42) = 49.99$, $p < .001$; $\eta^2_p = 0.70$; $F_{Critical\ thinkings}(2, 42) = 47.95$, $p < .001$; $\eta^2_p = .70$) and post-hoc tests (Scheffe) showed that all patterns differed from one another, except for *Results* where only the low-attention-pattern is significantly lower than in the results pattern. The Kruskal-Wallis test revealed that neither of the two other learning outcomes differed in the three patterns ($\chi^2_{Underpinnings} = 5.32$, df = 2, $p = .07$; $\chi^2_{Curiosity} = 3.98$, df = 2, $p = .14$).

Programme profiles with the *low-attention pattern* ($n = 21$), devote limited attention to the RR-learning outcomes in their course descriptions. The low average scores of each of the learning outcomes suggest that in a majority of modules the learning outcomes are not - or only vaguely - mentioned (therefore scoring 0 or 1). For programmes with the low-attention pattern, RR-learning outcomes seem not to be part of the main objectives of the programme.

Table 3-3

Descriptive Statistics of the Three Patterns (Low-attention pattern n = 21, Results pattern n = 16, Critical-thinking pattern n = 8)

Pattern	*M* (*SD*)	95% Confidence interval
Results		
Low-attention pattern	0.527 (0.084)	0.357 to 0.697
Results pattern	1.220 (0.096)	1.026 to 1.414
Critical-thinking pattern	0.840 (0.136)	0.565 to 1.115
Underpinnings		
Low-attention pattern	0.072 (0.022)	0.027 to 0.117
Results pattern	0.016 (0.025)	-0.035 to 0.068
Critical-thinking pattern	0.015 (0.036)	-0.058 to 0.088
Practical research skills		
Low-attention pattern	0.583 (0.056)	0.469 to 0.697
Results pattern	1.062 (0.065)	0.931 to 1.192
Critical-thinking pattern	1.545 (0.091)	1.360 to 1.730
Competence to be a researcher		
Low-attention pattern	0.268 (0.053)	0.162 to 0.374
Results pattern	0.776 (0.060)	0.654 to 0.897
Critical-thinking pattern	1.212 (0.085)	1.041 to 1.384
Critical thinking		
Low-attention pattern	0.567 (0.061)	0.444 to 0.691
Results pattern	0.984 (0.070)	0.843 to 1.126
Critical-thinking pattern	1.700 (0.099)	1.500 to 1.900
Curiosity		
Low-attention pattern	0.059 (0.044)	-0.030 to 0.147
Results pattern	0.245 (0.050)	0.144 to 0.346
Critical-thinking pattern	0.054 (0.071)	-0.089 to 0.197

Programmes of the *results pattern* ($n = 16$) devote, in their course descriptions, a relatively large amount of attention to acquiring knowledge of research results (*Results*). It is the most important RR-learning outcome. Given the high mean score, this learning outcome is mentioned in numerous modules of those programmes. The attention is significantly higher than in the low-attention pattern. The development of *Critical thinking*, *Practical research skills* and the *Competence to be a researcher* are mentioned as course learning outcomes, and are of approximately equal importance.

The level of attention to these three learning outcomes is situated in-between the two other patterns.

Programmes with a profile that fits in the *critical-thinking pattern* (n = 8) devote attention to four RR-learning outcomes in their course descriptions. There is a distinct attention paid to *Critical thinking* and to a lower extent, though still considerable, to *Practical research skills* and the *Competence to be a researcher*. Given the high average score of these three variables, there is an obvious indication of attention to these learning outcomes for a majority of the modules of programmes with this pattern. Gaining knowledge of research results (*Results*) is mentioned in the course learning outcomes, but it is not as pronounced as for the three previously mentioned learning outcomes, and there is much variation in the devoted attention to this learning outcome between the different programmes with this profile (see Table 3-3). What is distinctive for programmes with this profile is the outspoken attention to the development of competences or skills related to conducting research.

3.3.2 Exploration of Relationship between Patterns and Type of Programme and Discipline

A chi-squared analysis revealed a significant difference in the prevalence of the patterns between the different types of programmes , $\chi 2$ = 13.506, df = 4, p = .009. The different types of programmes have a different dominant pattern (Table 3-4). Most professional bachelor's programmes (15 out of 22) have a profile that fits with the low-attention pattern. For the academic bachelor's programmes, the low-attention pattern is also the most frequent. For the master's, most programmes have a profile following the results pattern (9 out of 14).

In addition, when looking at the relative frequencies, it appears that the professional bachelor's programmes are overrepresented in the low-attention pattern (71% of the programmes with a low-attention pattern are professional bachelor's programmes, whereas the professional bachelor's programmes represent 49% of the total sample).

Master's programmes are overrepresented in the other two patterns (56% of the programmes with a results pattern and 50% of the programmes with a critical-thinking pattern are master's programmes, whereas master's programmes account for only 31% of the total sample).

Table 3-4

The Distribution over the Patterns by Type of Programme (Proportion)

	Low-attention pattern	Results pattern	Critical-thinking pattern	Total
Professional bachelor	15 (0.71)	4 (0.25)	3 (0.38)	22 (0.49)
Academic bachelor	5 (0.24)	3 (0.19)	1 (0.13)	9 (0.20)
Master	1 (0.05)	9 (0.56)	4 (0.50)	14 (0.31)
Total	21 (1)	16 (1)	8 (1)	45 (1)

A chi-squared analysis showed that the distribution throughout the different patterns does not differ between the hard and the soft sciences, $\chi 2$ = 2.632, df = 2, $p <$ .268. The majority of the programmes in the soft sciences follow the low-attention pattern (9 out of 14)(see Table 3-5). In the hard sciences the low-attention pattern and the results pattern are roughly equally common (12 and 13 programmes respectively).

Table 3-5

The Distribution over the Patterns by Discipline (Proportion)

	Low-attention pattern	Results pattern	Critical-thinking pattern	Total
Hard	12 (0.57)	13 (0.81)	6 (0.75)	31 (0.69)
Soft	9 (0.43)	3 (0.19)	2 (0.25)	14 (0.31)
Total	21 (1)	16 (1)	8 (1)	45 (1)

3.3.3 Patterns of Programmes Leading to the Same Degree

Table 3-6 shows that programmes leading to the same degree can have different patterns. The biomedical sciences programmes are spread between the low-attention pattern and the results pattern. Two of the three programmes offering the same science degree have a low-attention pattern, the other has a critical-thinking pattern. The three academic programmes in applied sciences all have a different pattern.

Table 3-6

The Distribution over the Patterns for Programmes Offering the Same Degree

	Low-attention pattern	Results pattern	Critical-thinking pattern	Total
Professional bachelor degree in biomedical sciences	6	4	0	10
Professional bachelor degree in sciences	2	0	1	3
Academic bachelor degree in applied sciences	1	1	1	3

3.4 Conclusion

This study aimed to shed light upon the extent to which research integration forms part of students' daily educational experiences. Research integration is measured by the frequency of RR-learning outcomes in course descriptions. Research integration is given a central role in defining higher education by administrators and teachers, as well as students (Durning & Jenkins, 2005; Neumann, 1994; Rowland, 1996; Leisyte et al., 2009). However, until now it has been unclear how often research is integrated into teaching (Hattie and Marsh, 1996). Other studies often take the research activities of the teacher as a starting point to look how it influences teaching. Because the constraints of daily practice could hinder the intentions to integrate research this study takes the teaching setting as the starting point.

Whilst the results are coloured by the local context of higher education in Flanders, the study contributes to the literature on the teaching-research nexus as it is one of the first to analyse the frequency of research integration in such a large number of modules and programmes; it surpasses case-based level of other studies (van der Rijst & Jacobi, 2009), and is able to show whether different patterns can be discerned.

In the study three distinct patterns were identified: a low-attention pattern with an overall low attention for the RR-learning outcomes; a results pattern with a higher attention for the RR-learning outcomes, especially for acquiring knowledge of research results; and a critical-thinking pattern with the main focus on *Critical thinking*, *Practical research skills* and the *Competence to be a researcher*. The critical-thinking pattern could be considered as the pattern most focused on undertaking research or becoming a researcher. Hence, the study reveals that students' experiences with RR-learning outcomes differ in intensity.

For programmes with the low-attention pattern (about half of the programmes in the sample) the experience is rather limited. For programmes with more experience with RR-learning outcomes, the focus of attention differs. This can be either more oriented towards research results (the results pattern), or more oriented towards *Critical thinking*, *Practical research skills* and the *Competence to be a researcher* (the critical-thinking pattern). Given the perceived importance of the relationship between teaching and research (e.g., Taylor, 2007), the frequency of the low-attention pattern provokes questions concerning the relationship between perceptions and practices; between the importance attached to research integration and the possibilities to put it into practice.

The study also shows that the attention devoted in the course descriptions to gaining insight into the methodological and theoretical *Underpinnings* of research and the development of *Curiosity* towards evolutions in the discipline is very low. This is a conspicuous finding, as one could expect that a critical attitude towards information, knowledge, and knowledge construction builds partly on knowledge about how research is conducted and about implications of the use of particular theoretical frameworks. The low attention to the development of *Curiosity* conflicts with findings from interview studies, where teachers stated that they integrated research into their classes in order to create enthusiasm within students for a discipline; to make them experience the thrill of finding something new or unexpected (Elen & Verburgh, 2008; Leisyte et al., 2009). Further research is needed to investigate this in detail.

When analysing this sample for differences with regard to programme types (professional bachelor's, academic bachelor's and master's), different dominant patterns according to programme type were found. For the professional bachelor's programmes the low-attention pattern is most dominant, for the master's the results pattern. For the academic bachelor's programmes, the low-attention pattern is dominant too, but this is less clear when compared to the two other types. The dominance of the low-attention pattern among the professional bachelor's programmes is in line with expectations as research integration is often not considered as important for professional bachelor's as for academic programmes. Nevertheless, 38% of the programmes with a critical-thinking pattern - the pattern with the highest attention for the RR-learning outcomes - are professional bachelor's programmes, and there is only one of the nine academic bachelor's programmes with this pattern. Apparently, being critical in combination with *Practical research skills* and the *Competence to be a researcher* is for some professional bachelor's programmes a prominent aim. For academic bachelor's programmes, this seems to be less the case. This result indicates that the traditional distinction between professional and academic programmes is not as clear-cut as the name would indicate. Some professional programmes may have an important *Critical thinking* or research-like component. This is consistent with the findings of Lepori and Kyvik (2010) who described an academic drift in European non-university institutions offering professional programmes. It might be that *Critical thinking*, *Practical research skills* and the *Competence to be a researcher* become important when students are being prepared for the labour market, irrespective of whether the programme is a professional bachelor's or a master's. It is also possible that these learning outcomes are not particularly important when students are being prepared for further study, as is the case for academic bachelor's programmes. However, further research, with an a-select sample is needed to investigate this possible explanation.

The analysis of differences between disciplines was not in line with our expectations that the RR-learning outcomes would be more prominent in the soft sciences (Coate et al., 2001; Robertson & Bond, 2001). In the sample there are no clear differences between disciplines. Because of the opportunistic nature of the sample and the underrepresentation of the soft sciences, interpretations are to be made cautiously. It may relate to difficulties in defining what "research" is. What constitutes research differs between academics (Brew, 2001), and there are indications that these conceptions differ between disciplines (Prosser, Martin, Trigwell, Ramsden, & Lueckenhausen, 2005). This may have implications for the wording employed in the course descriptions.

The comparison between the programmes leading to the same degree indicates that these programmes pursue the RR-learning outcomes to a different extent. This evokes questions in regard to the comparability of degrees granted by different institutions. In Flanders, where there is no hierarchy or benchmarking between institutions, this finding urges further research.

When interpreting the results, attention should be paid to at least two limitations of this study. First, the results are based on an opportunistic sample. All programmes volunteered to participate, and were interested in the teaching-research nexus. This could bias the sample, and, more precisely, result in an overestimation of the importance of RR-learning outcomes. Nevertheless, approximately 13% of all regular Dutch programmes of the KU Leuven association are included. Second, the results are limited to the formal curriculum, which does not always mirror the enacted curriculum (Marsh & Willis, 1999). If certain things are not written down, it does not necessarily mean that they are not aimed at by the teacher, and vice versa, it is not the case that because something is written down, it is actually taught. A study of Visser-Wijnveen (2009) showed that the educational aims stated by teachers are more modest than what students report to have learned, especially in relation to awareness of research and research disposition. While awareness of research relates to the intangible nexus of Neumann (1992), and is not the focus of this study, the development of a research disposition is. Based on the study of Visser-Wijnveen (2009), it is possible that students perceive that they developed their *Competence to be a researcher* through studying a particular course, even when it is not stated as an explicit learning outcome of that course.

The study sheds light on a particular aspect of research-integration. It shows that research integration, in terms of RR-learning outcomes is, to a differing extent, part of students' daily experiences. At the same time, it immediately raises questions and ideas for further research. First of all, additional research with a random sample of programmes is needed in order to verify whether the same profiles can be discerned, to investigate the overall frequency of these profiles, and to study their relationship with programme types and discipline at a more profound level.
In addition, an interview study with teachers about what they aim to achieve through their modules, and how they evaluate students on these aims, would complement the method used in this study. Such an interview study could highlight similarities and discrepancies between the course descriptions and actual teaching and evaluation practices. In the interviews, specific attention should be paid to the importance of *Underpinnings* and the development of *Curiosity*.

Such an interview study is not feasible for a large number of programmes, due to the number of modules involved in a single programme. Therefore, a combination of both approaches may be the most promising approach.

Chapter 4: RI-approaches: Prevalence and Differences Between Modules[7]

Research integration into teaching is considered as a desirable characteristic of higher education. Despite pleas to strengthen research integration, not much is known about its current prevalence. This article reports on an study aimed at (1) identifying the prevalence of eight different research integration approaches within the first year of the undergraduate curriculum and (2) identifying module types combining specific research integration approaches. The study revealed that the most frequent research integration approach is so called Facts , an approach in which students are confronted with research results as if they were facts. Students' involvement in research-like activities is almost exclusively relevant for their own learning without being functional for the development of the discipline. **Insightful practice** is the most common module type, combining five different research integration approaches, which are Facts , Scientific facts, Research-based facts, Research methods and Segments- outcomes relevant. Practical implications and directions for further study are discussed.

4.1 Introduction

The relationship between research and teaching is considered as a core characteristic of higher education (Barnett, 1990; Clark, 1997). Most commonly, the relationship is looked at from the importance of research for teaching and learning and not from the importance of teaching for research (Roberston, 2007; Taylor, 2007).

[7] This chapter is submitted for publication [Verburgh, A., Schouteden, W. & Elen, J. (2013b). The prevalence of research integration approaches in the undergraduate curriculum. Manuscript under review.].

First year students consider the engagement with research as a defining element in university learning as compared to learning at secondary level (Levy & Petrulis, 2012). Teachers and administrators consider research and teaching as two intertwined fundaments of what constitutes higher education (Rowland, 1996). Pleas for enhancing research integration into teaching are common (e.g., Brew, 2006).

4.1.1 **Prevalence**

While many studies indicate that students are confronted with research during their studies (Levy & Petrulis, 2012; Neumann, 1992; Sin, 2012; van der Rijst & Jacobi, 2010; Verburgh et al., 2013a; Visser-Wijnveen et al. 2012; Zimbardi & Myatt, 2012), the prevalence of actual teaching practices in relation to research remains largely undisclosed (Hattie & Marsh, 1996; Verburgh et al., 2007). Verburgh et al. (2013c) found three different patterns in the prevalence of research-related learning outcomes, one in which research-related learning outcomes are hardly important, one in which critical thinking is important and one in which gaining insight into research results is the main focus. Sin (2012) investigated the place of research in the master's curriculum in three countries and found that in general the development of research skills is an important learning outcome while there are major differences between countries with respect to the intended learning outcomes of students' research participation. Van der Rijst and Jacobi (2010) scanned three bachelor's programmes at one university on their research integration practices. They found that students are taught about research results and research methods and that student are very often involved in research-like activities, such as research assignments. However student do seldom authentic research, i.e. research oriented to the development of knowledge and insights new to the discipline.

Gaining insight in the prevalence of actual teaching practices concerning research integration is particularly important as several scholars highlight the absence of an automatic relationship between teaching and research; it has to be deliberately established (e.g., Elton, 2001, Jenkins, Breen, Lindsay, & Brew, 2003; Westergaard, 1991). A seminal review of Hattie and Marsh (1996) showed that the empirical grounds for close relationships between teaching and research are limited. Research output and teaching satisfaction have a slightly positive, but close to zero, correlation. This implies that being a good researcher does not imply being a good teacher (Hattie & Marsh, 1996; Terenzini & Pascarella, 1994). Although research integration is considered fundamental in higher education (Clark, 1997), it is possible that the constraints of daily practice drive teaching and research apart (Leisyte et al., 2009) and hence hinder teachers to integrate research into teaching to the extent they desire (Neumann, 1992). Therefore, this study investigates the prevalence of research integration in the first year of the undergraduate curriculum. Based on the plea of Healey and Jenkins (2009) for research opportunities for all students, there is a deliberate focus on the experiences of all students.

4.1.2 **Typologies**

In order to investigate the prevalence of research integration, a tool to study research integration is needed. However, despite the agreement on the importance or value of research integration for students, consensus is lacking on the precise meaning of what the integration of research into teaching implies (Annala & Makinen, 2011). There have been several studies on how research can be integrated into the curriculum or how research experiences in the curriculum can be defined (Verburgh, et al., 2007). These studies resulted in different typologies of research integration, accompanied with a specific vocabulary (e.g., Healey, 2005; Neumann, 1992; Trowler & Wareham, 2008; Visser-Wijnveen et al., 2012; Zamorski, 2002; Zimbardi & Myatt, 2012). The models of Neumann and of Healey are probably best known.

Neumann (1992; 1994) discerned between three basic forms of research integration. She called them the tangible, intangible and global nexus. The tangible nexus refers to the use of scientific research results during classes. The intangible nexus is more related to the research process and to the development of research skills and attitudes of students. While both the tangible and intangible nexus refer to an integration of research into the teaching at the level of the individual teacher, the global nexus pertains to a connection at the departmental level, for example "between the total research involvement of the department and the teaching activity of that department" (Neumann, 1992, p. 166). A department's research activity, for instance, influences the type of specialised courses that are offered, or the orientation of the department's basic courses (Neumann, 1992).

Healey (2005; Healey & Jenkins, 2009) maked a distinction between four different types of research integration, placed on two axes (see Figure 4-1). One axe concerns whether students are considered as audience or as participants. The other axe concerns the design of the curriculum going from an emphasis on research content to an emphasis on research processes and research problems. According to Healey research integration into teaching can be: *research-led*, *research-oriented, research-based* and *research-tutored*. When teaching is *research-led*, the curriculum is constructed around subject content and teaching is based on the transmission model. When it is *research-oriented*, the curriculum is constructed to understand the processes of knowledge production as well as to learn that knowledge. In teaching special attention is devoted to learning inquiry skills. In *research-based teaching*, the curriculum is constructed around inquiry-based activities and the divisions between teachers and students are minimised because the high similarity between staff activities and student learning activities in the processes of inquiry. If teaching is *research-tutored* the curriculum is focused on students discussing and writing about research results.

Students are participants

	Research tutored Engaging in research discussions	Research-based Undertaking research and inquiry	
Emphasis on research content			Emphasis on research processes and problems
	Research-led Learning about current research in the discipline	Research-oriented Developing research and inquiry skills and techniques	

Students frequently are audience

Figure 4-1 Model of Healey and Jenkins (2009, p. 70)

Zamorski (2002, p. 422-423) developed a list of five ways of research integration,

- “Gaining knowledge from recent research (their teachers’ or others’) through taught courses and units.
- Understanding more fully the complex and provisional relationships between research and knowledge.
- The gradual development of various research skills during the passage of an undergraduate course.
- Learning about research methods and skills on specified units, such as a research methodology course or unit.
- Engaging in research activity or a research project as an integral part of a course or unit.”

In a study on ideal teaching-research relations, Visser-Wijnveen and her colleagues (2010, 2012) make, among other things, a list of four different approaches. The first is *learning about research*, containing lecturing and literature reading. The second is *inquiry learning,* containing analysis, assignments, discussions and reporting. Here students learn in a research-like way without doing actual research. A third approach is *simulation* which can be an individual or group research project, and the last approach is *participation*, in the teacher's own research or in the academic world.

The typologies have generally spoken three aspects in common. First, they identify different ways of teaching in which students come in contact with research. Or in other words, there are different approaches of research integration into teaching. For example, students perform research activities such as research discussions. In this contribution, teaching approaches that bring students in contact with research or research-like activities, are referred to as research integration approaches (RI-approaches).

Second, the typologies identify that teachers have particular learning outcomes in mind when using research integration. For example students are expected to learn about research methods, or there is a focus on learning disciplinary content. So sometimes, it is the intended learning outcome which makes that the learning environment is considered as a manifestation of research integration. Hence, research integration not only concerns RI-approaches, but also what students have to learn in relation to research, research-related learning outcomes.

And third, the typologies combine RI-approaches and research-related learning outcomes or they use them interchangeably (Elen & Verburgh, 2008). The tangible nexus (Neumann, 1992) for instance refers to a RI-approach, namely using research results in class, while the intangible nexus is more related to research-related learning outcomes, like developing research skills or specific research attitudes.

4.1.3 **Limitations**

While these typologies show a broad spectrum of possible ways to integrate research into teaching, they are difficult to use to investigate frequencies of RI-approaches, with the exception of the approaches of Visser-Wijnveen et al. (2012). Pure forms of the prototypes might be less common than the mixed ones and some typologies are not mutually exclusive. If combinations are common, too many practices will end up as "mixed" forms to be informative or illuminating. Sometimes one example of a research integration practice can be placed on two places within the same typology, while they are proposed to be distinct types of research integration. It is for example conceivable that a teacher aims at developing an advanced understanding of the complex relationship between research and knowledge through the engagement in research , which would be in the model of Zamorski (2002) two different ways of research integration (namely the second and the fifth). An indication of the difference between the prototypes and the actual practices in modules can be found in the study of Zimbardi and Myatt (2012) as they have a mixed model, in which different other models are combined, containing 25% of their examples. And also Visser-Wijnveen et al. (2012) studied modules, which are smaller units within a programme. They identified five module types, each combining different RI-approaches. Hence, an analytic, fine-grained tool is needed, in order to study the prevalence of RI-approaches and the nature of the combinations within modules.

Moreover, the situation in university colleges is under-investigated as most studies concentrate on universities with research-active academics. University colleges did originally not have a research mandate (Griffioen, de Jong, & Jak, 2013). Teaching was the main focus of teachers, and programs focused on the employability of graduates and partnership with employers (Prokou, 2008). Due to the implementation of the Bologna declaration, there is an increasing expectation for teachers to do research and to integrate research into their teaching in all types of higher education (Witte et al., 2008). There is a drift within these programs towards developing research as an ordinary activity alongside teaching, referred to as "academic drift" (Burgess, 1973) or "research drift" (Kyvik & Skodvin, 2003).

Therefore the existing typologies and research findings based on university environments might not be fully applicable to university colleges, because of the different position of research between these two types of institutions.

4.1.4 Aims

This study investigates the prevalence of RI-approaches in the first year of the undergraduate curriculum in university colleges. Therefore, an analytic tool to investigate the RI-approaches will first be explained. Next, the study will look at the prevalence of the different RI-approaches in the first year of undergraduate education. And lastly, it will study whether within one module, single RI-approaches or combinations of different RI-approaches are common. The study hence tries to identify module types in which the same single RI-approaches or combinations of RI-approaches are used. The study aims to feed the discussion on the research-teaching nexus, which is reproached to be rhetorically strong but empirically light (Coate et al., 2001).

4.2 Method

4.2.1 RI-approaches

In this study, RI-approaches are teaching approaches in which students come in contact with research or research-like activities. Based on the work of Elen et al. (2011), a scheme with two dimensions was developed: research outcomes and research processes. While others (Elsen et al., 2009; Healey & Jenkins, 2009) used research results and the research process as opposite positions on one dimension, they are here used as two independent dimensions. This will contribute to distinguish between different RI-approaches and to avoid overlap between approaches.

For the research processes, there are six possibilities: 1. No focus on the research background, 2. Research background is mentioned, 3. Research background is explained, 4. Research methods are explained, 5. Segments of a research process are run through; and 6. A full research cycle is made.

For the research outcomes there are two possibilities: research outcomes are relevant for the student or functional for the discipline. The dimension of research outcomes relates to the tension of originality (Beckman & Hensel, 2009; Elsen et al., 2009; Levy & Petrulis, 2012). This tension concerns the question to whom the outcomes should be original, to the student or to the discipline. For example, repeating a classic experiment leads to results that are new for the students but widely known in the discipline (or else they are original to the student but not original to the discipline). Here these activities are considered as "relevant for the student", they are not intended to add something to the discipline but the activity is oriented towards student learning. When a research activity is functional for the discipline, it implies that it aims at contributing to the knowledgebase of the discipline. When the research outcomes are functional for the discipline, this does not exclude that they are relevant for the student.

Combining the two dimensions, there are eight meaningful positions concerning research integration (Table 4-1). The dimension of research outcomes only becomes relevant when students are involved in doing (segments of) research. For the other possibilities the outcomes are functional for the discipline otherwise they would not be taught to the students.

Table 4-1

RI-approaches by Dimensions "Research Processes" and "Research Outcomes"

RI-approach	Research Processes	Research Outcomes
Facts	No focus on the research background	/
Scientific facts	Research background is mentioned	/
Research-based facts	Research background is explained	/
Research methods	Research methods are explained	/
Segments- outcomes relevant	Segments of research are run through	Relevant for students
Segments- outcomes functional		Functional for discipline
Full study- outcomes relevant	A full research cycle is made	Relevant for students
Full study- outcomes functional		Functional for discipline

Facts

In the RI-approach Facts students are confronted with descriptions of research results as if they were facts. The research background of the results is not mentioned. This approach can be experienced in various settings, for example, attending a traditional lecture, listening to an explanation of a theory by means of examples, participating in a Socratic conversation with a teacher posing questions and students answering, watching a documentary film, reading a text or explaining a given text to peers. The absence of any mentioning of the research background of the findings is the key feature for Facts.

Scientific facts

In this RI-approach students are confronted with research results. Here the research background is touched but no further explanation of the research is given. This approach can be experienced in similar settings as Facts, but with touching the research background. This can for example be that during the lecture or the documentary film the speaker says that research has revealed the findings or that the name of a researcher or the scientist is dropped, or with references in a text grounding the statements made. Distinctive for Scientific facts is the mentioning of the research background without further explanation.

Research-based facts

When students are confronted with research results and their background, this is labelled as Research-based facts. Explaining the results of a study and why this is relevant for the theory under study is an example of this approach. Other examples could be a teacher showing a documentary film exemplifying the experiment underlying the theory being studied or a teacher explaining her students why the value of a particular parameter changes year after year because of evolutions in the discipline. In this approach students are confronted with the results of research and the research behind those results.

Research methods

In this RI-approach students are confronted with research methods. For example, they have to read an explanation about an analysis technique and how it should be used or a teacher demonstrates how to handle a machine to do measurements. The focus is on explanations of a research method.

Segments- outcomes relevant

In the RI-approach Segments- outcomes relevant students are actively involved in doing one or more research steps. The research activity is not oriented towards advancing the discipline but it is oriented towards the learning of the student, it is relevant for the learning of the student. This approach can be experienced in diverse settings, e.g., during lab-sessions students have to repeat a classic experiment, students receive a question and they have to think about a method to investigate the question, students have to look up information about a given topic, or students have to select and apply an appropriate analysis technique to answer a given question. Students are actively involved in doing segments of research, and the activity is oriented towards student learning.

<u>Segments- outcomes functional</u>

Similar to the previous RI-approach students are involved in doing one or more research steps. But in contrast the results are here not only relevant for the students but also intended to add to the advancement of the discipline. For example, students have to fine-tune a distillation machine, so that the teacher can use it in her experiments, students have to administer questionnaires of a research study of the teacher, or students have to perform a statistical analysis on real research data. Students are actively involved in doing segments of research, and their activity is intended to be functional for the discipline.

<u>Full study- outcomes relevant</u>

In this RI-approach students are involved in a complete research cycle and the results of the study are relevant for the students. For example students have to do a small literature review on a chosen topic but the results of the review are not important or are already known by the teacher. Students' research is oriented towards student learning.

<u>Full study- outcomes functional</u>

In this RI-approach students are also involved in a complete research cycle, but in contrast to the previous approach, the results are intended to advance the discipline. Here students become members of the research community. For example students have to do a literature review that helps the teacher to develop a hypothesis for future experiments. In this approach the research students are involved in, is oriented towards the advancement of the discipline.

4.2.2 Participants

In this study 34 teachers from four bachelor's degree programmes participated in four different university colleges (Table 4-2). In the setting for this study, Flanders, Belgium, programmes are defined as structural entities leading to a degree, e.g., the bachelor in geography. Modules are smaller units within a degree programme. Within a programme, students have to follow several modules. It is pre-specified which modules students have to follow and between which modules they can choose. Upon registration at an institution, students select a specific programme. Organised programmes are organized in "phases" of 60 credits, comparable to one full year of study. In principle, first year students follow the first phase. To graduate from a bachelor's degree programme students have to gain 180 credits.

Table 4-2
Number of Discussed Modules

Programme	N
A	12
B	10
C	10
D	14

Because research integration might differ between disciplines (Coate, 2001; Colbeck, 1998; Robertson, 2001; Rowland, 1996), all programmes are selected within hard sciences (Biglan, 1973) in order to have comparable programmes. Three programmes have an academic orientation, one programme has a professional orientation. Programmes with an academic orientation are more theory-oriented, or oriented towards a more general, abstract introduction to a scientific discipline (Witte et al., 2008; Huisman & Kaiser, 2001). They are scaled at 5A level in the ISCED-97 classification system (International standard classification of education of the OECD). Programmes with a professional orientation are intended to prepare students for a particular occupation or profession which requires an extended study in higher education (Huisman & Kaiser, 2001). They are scaled at 5B medium level in the ISCED-97 classification system. The three programmes with an academic orientation lead to a similar degree. Besides their academic orientation they have tight links with the professional field.

4.2.3 **Interviews**

The teachers responsible for each first phase module participated in an individual semi-structured interview. The interviews lasted between 30 minutes to one hour. When teachers were responsible for more than one module, the different modules were discussed during the interview.

Teachers were asked to describe what happens in their classes, to talk about what they do during their contact moments with their students, what they expect from their students during and in between the contact moments, what they aim at or what they want their students to be able to do after finishing the module. Because teachers interpret research integration in a specific way (Neumann, 1992; Zamorski, 2002), there was no specific focus on research integration in the opening question of the interview. This more open approach was adopted in order to avoid that teachers would only talk about what they consider as research integration and that might be not as broad as it was conceptualised in this study. When needed, the interviewer asked clarification questions, in order to gain enough information about research integration. For example when a teacher told that she explained a theory to her students during a lecture, the interviewer asked how she did it, and whether she referred to the researcher or the research underlying the theory. The interviewer made sure that the RI-approaches in the module were clear.

4.2.4 **Analyses**

All interviews were audio taped and transcribed verbatim. The unit of analysis is the module. For each module the authors identified which RI-approaches were used in the module (see Table 4-1). So, for each module, there were eight scores, one score for each of the eight RI-approaches.

For the identification of the prevalence of the different RI-approaches a detailed and analytical scoring protocol was developed during an iterative process (Miles & Huberman, 1994). Based on an initial scoring protocol the two first authors scored individually two interviews, discussed their analysis together and refined the scoring protocol and scored again two other interviews. After three rounds the protocol was detailed enough to score reliably, so that at most one score per module was different. All remaining interviews were scored by the first author. The second author scored two additional interviews to guarantee consistency in the scoring.

The scoring of the RI-approaches evoked a continuous deliberation. On the one hand there was the intention to be open-minded to understand what is going on. On the other hand there was the possible hazard of making "research integration" an umbrella-concept, adding nothing to our understanding of research integration. The difficulty was predominant in making a distinction between *Segments- outcomes relevant* and exercises. In order to be considered as *Segments- outcomes relevant* students had to do an activity that was closely related to research or that in its original form was considered as research (e.g., the replication of a classic experiment is considered as *Segments- outcomes relevant* and solving an equation not). Moreover, when something is not coded as a RI-approach, it does not imply that the approach is less valuable than an approach labelled as "research integration".

For the description of module types combining RI-approaches, the original transcripts were read again, per module type. This was considered necessary in order to adequately describe the sequences within the module and the intentions of the teachers, after the analytic coding of the approaches.

4.3 Results

4.3.1 The Prevalence of RI-approaches

The most prominent approach of research integration in the first year was an approach in which students were confronted with research results, without the research background being mentioned (Facts) (see Table 4-3). In almost 90% percent of all modules this approach was used. In about half of the first year modules the research background of information was mentioned, without further elaboration (Scientific facts). In about 40% of the modules the research background was discussed (Research-based facts). In about four out of ten modules research methods were explained (Research methods). In almost 60% of the modules students did segments of research, in which the actual outcomes of that research were not oriented towards the advancement of the discipline but relevant for student learning (Segments- outcomes relevant). In one module students had to do segments of a full study of which the results were functional for the advancement of the discipline (Segments or Full study- outcomes functional) and in one module students had to complete an entire research study (Full study- outcomes functional). In none of the modules students had to do a complete study of which the results were solely oriented towards student learning (Full study- outcomes relevant).

Table 4-3

Proportion of Modules Using a Particular RI-approach (N = 46)

RI-approaches	Proportion of modules
Facts	0.87
Scientific facts	0.52
Research-based facts	0.39
Research methods	0.43
Segments- outcomes relevant	0.57
Segments- outcomes functional	0.02
Full study- outcomes relevant	0.00
Full study- outcomes functional	0.02

4.3.2 **Modules Types Combining RI-approaches**

Next, we looked for module types in which single RI-approaches or combinations of RI-approaches were used within the same module. They were labelled as " module types" when they occurred three times or more. Hence, the module types are not based on theoretical or logical combinations but on what occurs in this specific sample. Three was chosen as a minimum in order to avoid too much particularities.

The analysis revealed six different module types, i.e. six combinations of one or more RI-approaches with three or more occurrences (see Table 4-4).

Table 4-4
Combinations of RI-approaches within one Module (N = 46)

	Modules Types																	
RI-approaches*	A	B	C	D	E	F	g	h	i	j	k	l	m	n	o	p	q	r
Facts	x	X	x	x	x	x		x	x				x	x	x	x	x	
Scientific facts	x		x		x		x				x		x	x		x	x	
Research-based facts	x					x		x				x	x		x		x	
Research methods	x			x							x	x		x	x	x		x
Segments- outcomes relevant	x			x	x	x			x	x						x	x	x
Segments- outcomes functional																		x
Full study- outcomes relevant																		
Full study- outcomes functional																		x
N	9	6	5	5	3	3	2	2	2	1	1	1	1	1	1	1	1	1

Note. A = **Insightful practice**, B = **Facts only**, C = **Facts plus**, D = **Research process explained and applied**, *E* = **Facts research-like illustrated**, F = **Research-based facts explained and illustrated**
*Module types with three or more occurrences are labelled

Facts only *(N = 6)*

In a module of the type **Facts only** the RI-approach Facts is used. This implies that student were confronted with research results as if they were facts, the research background of the findings was not mentioned.

Facts plus *(N = 5)*

A module of the type **Facts plus** contained two RI-approaches: Facts and Scientific facts. Hence, in this type students were confronted with research results of which sometimes the research background was not mentioned (Facts) and sometimes mentioned (Scientific facts).

Facts research-like illustrated *(N = 3)*

Modules of the type **Facts research-like illustrated** combined three RI-approaches: Facts, Scientific facts and Segments- outcomes relevant. The three modules of this type all consisted of two parts, one part in which students were confronted with research results as Facts and Scientific facts (so, similar to the type **Facts plus**), and the other part, in which students had to do a kind of cookbook labs, in which they had to follow a prescribed procedure, as a kind of illustration of the theory discussed. The order of the two parts could alter.

Research-based facts explained and illustrated *(N = 3)*

Also in this type, modules contained three RI-approaches: Facts, Research-based facts, and Segments- outcomes relevant, also divided in two parts. In one part the teacher confronted students with research results, sometimes these results were presented as if it were facts (Facts), sometimes the research background was discussed in depth (Research-based facts). In the other part of the module students were involved in doing segments of research (Segments- outcomes relevant). These activities are considered as illustrations of the theory under discussion.

Research process explained and applied *(N = 5)*

Modules of this type combined three RI-approaches: Facts, Research methods, and Segments- outcomes relevant. Here the modules existed mostly of a more theoretical part in which the content was discussed as facts (Facts) and a practical part. In the practical parts students received explanations about specific applications or measurement instruments (Research methods) and afterwards, they had to use these instruments to do some measurements (Segments- outcomes relevant). Here the involvement in research-like activities is considered as an application of the theory under discussion.

Insightful practice *(N = 9)*

A module of the type **Insightful practice** used five RI-approaches: Facts, Scientific facts, Research-based facts, Research methods and Segments- outcomes relevant. Generally, a module of this type consisted of two parts, one part in which students were confronted with research results, in which research background were sometimes not mentioned (Facts), sometimes touched (Scientific facts) and sometimes explained into more detail (Research-based facts). In the other part there was an explanation of specific research methods (Research methods) and students had to do segments of research of which the results are relevant for students (Segments- outcomes relevant). Sometimes this second part also confronted students with research results as facts (Facts). Hence, the second part is similar to **Research process explained and applied**. Here, the involvement in research-like activities was considered as an insightful application of the theory.

Modules only using one single approach of research integration were a minority (9 out of 46 modules).

4.4 Discussion

This study investigated the prevalence of RI-approaches in the first year of the undergraduate curriculum. The study revealed that a RI-approach in which students are confronted with research results, without mentioning their research background (Facts), is the most prominent approach. In nearly 60% of the modules students undertake research activities for which the results are relevant for the students but not for the advancement of the discipline (Segments- outcomes relevant).

The study revealed that in the majority of modules different RI-approaches are combined (37/46). The type **Insightful practice** is the most common (N = 9). In daily practice single RI-approaches within one module are a minority. This study points out, similar to the findings of Visser et al. (2012), that the commonly used typologies, such as the one of Healey (2005; Healey & Jenkins, 2009) are not really suited to label modules as a whole, but more for specific parts of those modules.

The study is one of the first to investigate the prevalence of RI-approaches in the undergraduate curriculum, and this in university colleges. Although research integration is a multifaceted concept, this study only concentrated on one aspect of research integration. By doing so, the study wanted to avoid the critiqued conceptual vagueness and slippery way of using concepts in the field of research integration (Trowler & Wareham, 2008). The results indicates that precise conceptualisations are indeed needed because the prevalence of the different RI-approaches differs and most probably not all approaches are equally valued. We concur with Spronken-Smith et al. (2012) that talking about research integration as a unified approach will miss nuances associated with different practices concerning research integration.

An inclusive position is taken in identifying the eight RI-approaches. First, teaching practices which involved research-like activities with results relevant for students, but not oriented towards advancement for the disciplines are included. This is a similar position as Brew (2010) and Levy and Petrulis (2012), but opposite to Beckman and Hensel (2009). Moreover, also approaches in which students are confronted with research without doing research(-like) activities are included. This is similar to "students as audience" of Healey & Jenkins (2009) and more inclusive than the model of Zimbardi and Myatt (2012). The RI-approaches show resemblance with the approaches of Visser et al. (2012), but they are more detailed. Learning about research is differentiated in Facts, Scientific facts and Research-based facts . Inquiry learning can be considered as Segments- outcomes relevant, Simulation as Segments or Full study- outcomes relevant, and Participation as Segments or Full study- outcomes functional. There is no approach as Research methods among the approaches of Visser et al. (2012).

Based on the results, one could conclude that research integration is part of the undergraduate curriculum in university colleges. But, it also evokes discussion about the desired way of linking research and teaching in higher education. When the claim is made that research integration is a distinctive characteristic of higher education, which sets it apart from other types of education like secondary education, it is disputable if one has a RI-approach as Facts in mind. The content of many other types of education, is also based on research and teachers will confront their students with that content. Hence, it is most probably that also in other types of education, Facts will be an RI-approach.

Furthermore, the active involvement in research-like activities, is almost exclusively concentrated on activities of which the findings of the activities are not the main point of doing the activity (Segments- outcomes relevant). There is no research for the sake of truth, but for the learning of the student. This fits with what Simons and Elen (2007) would define as the functional approach to research integration. This approach departs from how research can be used in education, to be functional for the development of skills valued in society. It is opposed to the idealistic approach, in which academic education is considered as participation in research. The idealistic approach builds on the edifying role of academic inquiry, which is characterised by the pursuit of knowledge and truth: the actual aim of research. The authors warn for ambivalences and confusion because in discussions both approaches often make reference to one another. In particular the functional approach often uses arguments from the idealistic approach (Simons & Elen, 2007). Based on the findings of this study, it can be concluded that in the first year of the undergraduate curriculum there is no idealistic approach of research integration. During the first year students are confronted with research and research-like activities in more modest RI-approaches and one could rather speak of a functional approach.

The programmes under investigation are all from university colleges. It might be that the prevalence of RI-approaches or the found module types are different in a university context. The findings of Visser et al. (2012) give an indication that the teachers own research activity might be more prominent. The findings of Van der Rijst and Jacobi (2010) show that also in a university setting students are seldom involved in authentic research. A comparison between universities and university colleges would be an interesting direction for future research.

The insight in the daily practice concerning research integration found in this study is limited in three ways. The first is that the analysis is based on interviews, on what teachers tell they do. The actual teaching practice is not investigated. There might be differences in teachers' descriptions of what they do and what they actually do (Mälkki & Lindblom-Ylänne, 2012). The second shortcoming is that the analysis is done at the level of the modules and there is no indication of the importance of the RI-approach within that module or within the programme. A RI-approach can be the core of a module or only a small aspect. Similarly, a module can be a capstone module in a programme or a module of minor importance. Lastly, the study looks at RI-approaches independent from the intended learning outcomes. It would be valuable for further research to investigate the relationship between RI-approaches and learning outcomes, as this link is often made in the typologies of research integration.

The study provides a precise vocabulary to talk about RI-approaches and a tool to study the prevalence of a broad range of different RI-approaches. Therefore it can feed the discussion about the aspired prevalence of desirable RI-approaches.

Chapter 5: Relationship between RR-learning Outcomes and RI-approaches in Modules[8]

Despite pleas to strengthen research integration in higher education, insight in practices of research integration is fragmented. This article focuses on the interplay between research integration approaches (RI-approaches) and research-related learning outcomes (RR-learning outcomes). Data were gathered through interviews with higher education teachers about their first year modules. This article investigates whether six specific combinations of RI-approaches, labelled as module types, aim at specific RR-learning outcomes. The analyses reveal some specific relations between the module types and learning outcomes, such as some module types aim more than others at gaining insight in the underpinnings of research. The module types do not differ on all RR-learning outcomes studied. Acquiring knowledge of research results and developing research skills, as RR-learning outcomes, do not distinguish module types from one another. However, which research skills are aimed at differs between the module types. The results also indicate that combining many RI-approaches does not imply that more RR-learning outcomes are aimed at. The study point at the value of fine-grained distinctions in RI-approaches and RR-learning outcomes, studied at module level, to improve the understanding of research integration and its possible benefits.

[8] This chapter is submitted for publication [Verburgh, Schouteden, & Elen, J. (2013a). Research integration in the first year: Relationships between research-related learning outcomes and research integration approaches. Manuscript submitted for publication.].

5.1 Introduction

5.1.1 **Importance of Research Integration**

Confronting students with research is considered important in higher education (Visser-Wijnveen et al., 2012). It is assumed to be conducive for student learning (Brown & Mc Cartney, 1998). The Boyer commission (1998) for instance proposed to stimulate student learning through inquiry learning and research opportunities. Similarly, the European University Association (2007) considers research-based education as a strength of European Universities and states that students need to gain research experience and develop research-related skills.

Insight in teaching practices concerning research integration is therefore valuable. It is particularly important as several scholars state that there is no automatic relation between teaching and research. It has to be deliberately established (e.g., Elton, 2001; Jenkins et al., 2003; Westergaard, 1991; Zubrick et al., 2001). It may be that the constraints of daily practice hinder teachers in integrating research into teaching to the extent or the way they desire (Leisyte et al., 2009). This study will focus on the nature of the teaching practices of research integration in a cross section of the curriculum.

5.1.2 **Categorisations of Research Integration**

Insight in practices of research integration is fragmented (Hattie & Marsh, 1996; Verburgh et al., 2007). In the literature different terminologies are used, e.g., inquiry learning, undergraduate research, research opportunities or research-based teaching, and there is a tendency to use terms negligently, which hampers a precise understanding (Spronken-Smith et al., 2012; Trowler & Wareham, 2008).

Several categorisations were proposed on how research can be integrated into the curriculum or how research experiences in the curriculum can be defined (e.g., Trowler & Wareham, 2008; Visser-Wijnveen et al., 2010; Visser-Wijnveen et al., 2012; Zamorski, 2002; Zimbardi & Myatt, 2012). When analysing those categorisations, at least two different facets of research integration become apparent: research-related learning outcomes and research integration approaches(Elen & Verburgh, 2008).

A first facet of research integration concerns learning outcomes in relation to research. Teachers intentionally integrate research into their teaching because they want their students to achieve specific learning outcomes (Elen & Verburgh, 2008). In this interpretation of research integration, it is the intended learning outcome that defines whether teaching is considered as a manifestation of research integration. Therefore these learning outcomes can be referred to as "research-related learning outcomes" (RR-learning outcomes). Different RR- learning outcomes can be aimed at. They can for example imply gaining knowledge from (recent) research (Durning & Jenkins, 2005; Neumann, 1992; Zamorski, 2002). But they can also refer to understanding the underpinnings of research (Durning & Jenkins, 2005; Visser-Wijnveen et al., 2010, 2012; Zamorski, 2002; Zimbardi & Myatt, 2012). Another possible research-related learning outcome is the development of research skills or of a research disposition (Durning & Jenkins, 2005; Neumann, 1992; Healey, 2005; Visser-Wijnveen et al., 2010, 2012; Zamorski, 2002; Zimbardi & Myatt, 2012).

A second facet of research integration concerns the approaches used to integrate research in the learning environment. In this interpretation, it is the teaching approach that defines whether research integration occurs. Therefore these approaches can be referred to as "research integration approaches" (RI-approaches). Different RI-approaches can be identified. One possible RI-approach is that students engage in a simulation of research or that students engage in an authentic research project intended to contribute to the discipline (e.g., Durning & Jenkins, 2005; Zamorski, 2002; Zimbardi & Myatt, 2012). In the model of Healey (2005) this would be labelled as research-tutored or research-based teaching depending on whether the activity is a simulation or an authentic research project. Another approach could be that students are confronted with research methods (Durning & Jenkins, 2005; Zamorksi, 2002). The confrontation of students with research results during lectures or assignments is still another possible RI-approach (e.g., Healey, 2005; Neumann, 1992; Zamorski, 2002). This approach fits with the tangible nexus of Neumann (1992).

It is important to note that there is discussion about the breadth of what counts as research integration. For example, Zimbardi and Myatt (2012) confine research integration to teaching approaches that imply active involvement of students in research-(like) activities. Others restrict it to authentic research activities where the aim is that students' work contributes to the knowledge in the discipline (Council on Undergraduate Research (CUR.org)).

Categorisations on research integration do not always clearly make the distinction between RI-approaches and RR-learning outcomes (Elen & Verburgh, 2008). In some categorisations RI-approaches and RR-learning outcomes are placed side by side. For example, research-tutored education of Healey (2005) refers to engaging in research discussions (an approach), while his research-led education refers to learning about current research results (a learning outcome). Categorisations often suggest that within one learning environment different categories of research integration can be applied successively. Nevertheless, within some categorisations a specific learning environment may fit simultaneously on two places of the categorisation. For example, when students have to do a small research study in order to understand the process of knowledge construction, this activity would fit under research-based teaching and the learning outcome under research-oriented teaching in the model van Healey (2005).

Making a clear distinction between the RR-learning outcomes and the RI-approaches could foster an understanding of the subtleties of research integration and the interplay between learning outcomes and approaches, for at least two reasons. First, the learning outcomes attributed to research, are more diverse than those included in the categorisations. Interview studies on the advantages and disadvantages of the relation between teaching and research indicate the development of research skills, the acquisition of knowledge of current research, the development of critical thinking, the ability to deal with complex problems and to motivate students or to make them curious as important RR-learning outcomes (e.g., Leisyte et al., 2009; Rowland, 1996; Elen & Verburgh, 2008). The categorisations most commonly are confined to research skills and knowlegde of research results.

Second, the same RR-learning outcome might be aimed at with different RI-approaches. Studies on the effect of undergraduate research on student learning show a wide range of learning outcomes. Undergraduate research is found beneficial for the development of research skills, personal development (e.g., self confidence or perseverance), thinking and working like a scientist, clarification and preparation of career choices, learning a topic in depth (Bauer & Bennett, 2003; Hunter et al., 2007; Lopatto, 2009). Farrand-Zimbardi, van der Burg and Myatt (2010) found that the intended learning outcomes of other research experiences than undergraduate research were comparable to the learning outcomes of undergraduate research. This raises questions about the specific relations between approaches and learning outcomes. A question that comes to the front is whether there are different combinations of approaches to achieve specific combinations of learning outcomes.

5.1.3 The Relation between Combinations of Approaches and Learning Outcomes

An indication of specific relations between combinations of approaches with learning outcomes was found by Visser-Wijnveen et al. (2010, 2012). Based on twelve interviews complemented with logbooks of faculty members about a module of their own choice, Visser-Wijnveen and her colleagues identified five distinct module types, each characterised by a specific combination of RI-approaches and RR-learning outcomes (see Table 5-1).

In type A, the intended learning is directed towards acquiring disciplinary content knowledge (academic knowledge) and the development of research skills (train researcher). The approaches used are lectures that students have to prepare for with assignments. Assignments are part of inquiry learning during which students learn in a research-like way without doing actual research. The modules in type B are intended to understand the discipline and being able to participate in a scholarly debate as part of an academic disposition, mainly understood here as being critical. The teachers require their students to read literature in preparation of class discussions. In these modules students also have to write a paper. Type C modules introduce students to the literature during lectures and discussions, followed by individual research projects of students. Type C modules have a double orientation towards the development of research skills in combination with either the acquisition of academic knowledge or an academic disposition. Modules of type D are similar to type C. The difference pertains to a closer connection to the work of the teacher and the replacement of lectures are by literature reading. Whereas in type C students are involved in research-like activities, in D and E they have to do the real work. In E they participate in the research of the teacher with the intention to develop research skills and academic knowledge.

Table 5-1

Description of the Module Types of Visser-Wijnveen et al. (2010, 2012)

Type	Description	Approach	Learning outcome
A	Using the teacher's own research to illustrate the subject matter	Learning about research (Lecturing) Inquiry learning (Assignments)	Academic knowledge Train researcher
B	Focussing on the researcher's disposition and position	Learning about research (literature reading) Inquiry learning (Discussions/ Reporting)	Academic disposition
C	Introducing students to literature, after which students conduct research projects	Learning about research (Lecturing) Inquiry learning (Discussions) Simulation (individual project)	Train researcher (academic disposition/ knowledge)
D	Follow in the teacher's footsteps	Learning about research (literature reading) Inquiry learning (Discussions) Simulation (individual project)	Train researcher Academic knowledge
E	Participation in the teacher's research	Participation	Train researcher Academic knowledge

5.1.4 **Aims**

The present study focuses on the interplay of RI-approaches and RR-learning outcomes within a curricular cross section of all first year modules of four programmes. The general purpose is to investigate whether specific combinations of RI-approaches correspond to specific combinations of RR-learning outcomes. It builds on the results of an analysis of the frequency of RI-approaches and identified module types as combinations of one or more RI-approaches (Verburgh et al., 2013b). The relation between module types and intended learning outcomes is studied.

Therefore first the frequency of RR-learning outcomes is analysed. Next an in-depth analysis of the relation between module types and RR-learning outcomes is made.

5.2 Method

5.2.1 Participants

In the study teachers of 46 first year modules participated in an individual semi-structured interview. Participants teach in four different programmes at four different university colleges. In the setting for this study, Belgium, programmes are structural entities leading to a degree, e.g., bachelor in geography. Modules are smaller units within a degree programme. Because research integration might differ between disciplines (Coate et al., 2001; Colbeck, 1998; Robertson & Bond, 2001; Rowland, 1996), all programmes participating in the study belong to the same disciplinary area: hard sciences (Biglan, 1973).

5.2.2 Interviews

Teachers were asked to describe what happens in their classes, what they do during the contact moments with their students, what they expect from them during and in between the contact moments and what they want their students to be able to do after finishing the module. Because teachers interpret research integration in a specific way (Neumann, 1992; Zamorski, 2002) there was no specific focus on research integration in the opening question of the interview. This more open approach was adopted to avoid that teachers would only talk about what they consider as research integration. The interviewer asked clarifying questions, in order to gain enough information. For example when a teacher told that she explained a theory to her students during a lecture, the interviewer asked how she did it, and whether the teacher referred to the researcher or the research underlying the theory. The interview continued until all RI-approaches and all intended RR-learning outcomes in the module were clear. Interviews lasted between thirty minutes to one hour.

5.2.3 Analyses

All interviews were audio taped and transcribed verbatim. They were first analysed concerning the prevalence of the identified RR-learning outcomes and RI-approaches. For each module the authors identified what RR-learning outcomes were aimed at based on what the teacher said she wanted her students to achieve. The researcher not the teacher decided if the learning outcome stated by the teacher was classified as a RR-learning outcome. The RR-learning outcomes were not communicated to the teachers to avoid misunderstandings and socially desirable answers. The authors also identified what RI-approaches were used in the module. Next, all interviews were compared and based on the combination of different RI-approaches used within one module, "module types" were identified (Verburgh et al., 2013b). A module type is a combination of one or more approaches within one module (see Table 5-3). To be identified as a module type, the specific combination of RI-approaches had to occur in at least three modules. This was the case for six combinations. Using these six module types 31 out of originally 46 modules could be described. The last step in the analysis was the identification of the relation between the RR-learning outcomes and the module types. The different module types were compared on the RR-learning outcomes. In that comparison only differentiating learning outcomes were included. So learning outcomes common to or absent in all modules were excluded. Because of the small sample no statistical tests were done.

For the analysis of the interviews a detailed protocol was developed during an iterative process (Miles & Huberman, 1994). Based on an initial protocol the two first authors analysed individually two interviews, discussed their analysis together and refined the protocol and analysed again two other interviews. After three iterations the protocol was detailed enough to analyse reliably, so that at most two codes per module were different. There was agreement on 93 percent of the codes. All other interviews were then analysed by the first author. The second author analysed two additional interviews to guarantee consistency in the analysis.

5.2.4 **Variables**

RR- learning outcomes

Seven basic RR-learning outcomes were looked at, based on Verburgh et al. (2013c). Learning outcomes refer to what teachers want their students to learn (Anderson & Krathwohl, 2001). The RR-learning outcomes will be italicised.

In *Results* students have to acquire knowledge, based on research findings. *Underpinnings* refers to the insight in the underlying theoretical or methodological underpinnings of research. Students should gain insight in those underpinnings and its consequences for the meaning of the findings. In *Practical research skills* a single research skill is aimed at, such as being able to report research findings or find literature. The *Competence to be a researcher* is more encompassing than *Practical research skills*. Here students should be able to add something new. *Critical thinking* refers to the attitude to question information, knowledge or arguments. *Curiosity* refers to being curious towards evolutions in the discipline. *Practice* concerns the ability of students to look for research results to solve a problem or, the other way around, to see the practical implications or usefulness of a research result in their (future) professional practice. It is more than mere application of the content taught, students have to be able to add something new or to integrate new information.

Practical research skills and *Critical thinking* are further refined, because when teachers state that they aim at developing students research skills or ability to think critically they have very different skills and attitudes in mind (Schouteden et al., 2012). *Practical research skills* can refer to skills related to different research steps (*Formulating a research question, Finding literature, Developing a design, Collecting data, Analysing data, Formulating a conclusion* and *Report*). Developing a *Research attitude* and a *Systematic way of reasoning* are also considered as *Practical research skills*.

Teachers' interpretations of critical thinking can be divided into four categories: *Critical thinking towards oneself*, the own way of acting and the own frame of reference; *Critical thinking towards information* and how it is developed; *Being conscious of the perspective of others;* and *Being able to handle uncertainty*.

RI-approaches

Taking a broad interpretation of what counts as research integration, a distinction between eight different RI-approaches (see Table 5-2) is made (see for details, Verburgh et al., 2013b). For sake of clarity, RI-approaches will be underlined in the text.

Table 5-2

Eight RI-approaches (Verburgh et al., 2013b)

RI-approach	Description
Facts	Students are confronted with descriptions of research results as if they were facts. The research background of the results is not mentioned. This approach can be experienced in various settings, for example, attending a lecture, listening to an explanation of a theory by means of examples, watching a documentary film, reading a text or explaining a given text to peers.
Scientific facts	Students are also confronted with research results. Here the research background is touched but no further explanation of the research is given. This approach can be experienced in similar settings as Facts, but with briefly mentioning the research background. This can for example be that during the lecture or the documentary film the speaker says that research has revealed these findings, that the name of a researcher or the scientist is dropped, or that references in a text ground the statements made.

Research-based facts	Students are confronted with research results and the research behind those results. Explaining the results of a study and why this is relevant for the theory under study is an example of this approach, as well as a teacher showing a documentary film exemplifying the experiment underlying the theory being studied.
Research methods	Students are confronted with research methods. For example, they receive an explanation about an analysis technique and how it should be used or a teacher demonstrating how to handle a machine to do measurements.
Segments-outcomes relevant	Students are actively involved in doing one or more research steps. The research activity is not oriented towards advancing the discipline but it is relevant for the learning of the student. This approach can be experienced in diverse settings, e.g., during lab-sessions students have to repeat a classic experiment, students receive a question and they have to think about a method to investigate the question or students have to look up information about a given topic.
Segments-outcomes functional	Students are involved in doing one or more research steps, as in Segments- outcomes relevant. The results are here however not only relevant for the students but also intended to add to the advancement of the discipline. They are functional for the discipline. For example, students have to fine-tune a distillation machine, so that the teacher can use it in her experiments, students have to administer questionnaires of a research study of the teacher, or students have to perform a statistical analysis on real research data.
Full study-outcomes relevant	Students are involved in a complete research cycle and the results of the study are relevant for the student. Students do for example a small literature review on a chosen topic but the results of the review are not important or are already known by the teacher.
Full study-outcomes functional	Students are also involved in a complete research cycle, but in contrast to the previous approach, the results are intended to advance the discipline. Here students become members of the research community. For example students have to do a literature review that helps the teacher to develop a hypothesis for a future experiments.

Module types

This study compares six module types (combinations of RI-approaches) based on Verburgh et al. (2013b) (see Table 5-3). The module types will be written in bold.

In **Facts only** (*N* = *6*) student are confronted with research results as if they were facts (Facts). The research background of the findings is not mentioned.

In modules of **Facts plus** *(N = 5)* students are confronted with research results of which sometimes the research background is not mentioned (Facts) and sometimes mentioned (Scientific facts).

Modules of **Facts research-like illustrated** *(N = 3)* consist of two parts, one part in which students are confronted with research results as Facts and Scientific facts, and the other part, in which students have to do a kind of cookbook labs, in which they have to follow a prescribed procedure, as a kind of illustration of the theory discussed. The order of the two parts may alter.

In **Research-based facts explained and illustrated** *(N = 3)* modules are divided in two parts, similar to **Facts research-like illustrated**. In one part the teacher confronts students with research results, sometimes these results are presented as if it were facts (Facts), sometimes the research background is discussed in depth (Research-based facts). In the other part of the module students are involved in doing segments of research (Segments- outcomes relevant). These activities are considered as illustrations of the theory under discussion.

Modules of **Research process explained and applied** *(N = 5)* exist mostly of a more theoretical part in which the content is discussed as facts (Facts) and a practical part. In the practical parts students receive explanations about specific applications or measurement instruments (Research methods) and afterwards, they have to use these instruments to do some measurements (Segments- outcomes relevant). Here the involvement in research-like activities is considered as an application of the theory under discussion.

A module of the type **Insightful practice** (N = 9) generally consists of two parts, one part in which students are confronted with research results, in which research background is sometimes not mentioned (Facts), sometimes touched

(Scientific facts) and sometimes explained into more detail (Research-based facts). In the other part there is an explanation of specific research methods (Research methods) and students have to do segments of research of which the results are relevant for students (Segments- outcomes relevant). Sometimes this second part also confronts students with research results as facts (Facts). Hence, the second part resembles **Research process explained and applied**. The involvement in research-like activities is considered as an insightful application of the theory.

Table 5-3

Module Types as Combinations of RI-approaches (N = 31)

	Module type					
RI-approaches used	**Facts Only**	**Facts plus**	**Facts research-like illustrated**	**RB-facts explained &**	**Research process explained and**	**Insightful practice**
Facts	X	X	X	X	X	X
Scientific facts		X	X			X
Research-based facts				X		X
Research methods					X	X
Segments-outcomes relevant			X	X	X	X
Segments-outcomes functional						
Full study-outcomes relevant						
Full study-outcomes functional						
N	6	5	3	3	5	9

5.3 Results

5.3.1 Prevalence of RR-learning Outcomes

The analysis of the interviews revealed that in all modules teachers want their students to acquire knowledge from research results (*Results*) (See Total column of Table 5-4). In one third of the modules understanding the underpinning of research is pursued (*Underpinnings*). It also shows that the development of *Practical research skills* is aimed at in all modules, whereas the *Competence to be a researcher* is not aimed at in these first year modules. The development of *Critical thinking* is an aim in two thirds of the modules. About 40 percent of the modules tries to develop students' *Curiosity* towards developments in the discipline. The application of research results to solve an unknown practical problem (*Practice*) is not an aim in the first year modules under study.

When looking in more detail at *Practical research skills*, the analysis shows that developing a *Systematic way of reasoning* is most common, with an occurrence of almost 70 percent. *Collecting data, Formulating a conclusion* and *Report* are each in about one third of the cases aimed at. Learning how to *Develop a design,* to *Analyse data*, and the development of a *Research attitude*, are each in about 10 to 15 percent of the cases strived for. *Formulating a research question* or *Finding literature* are almost never intended.

The detailed analysis of *Critical thinking* reveals that teachers, when they aim at *Critical thinking*, predominantly want their students to be able to think critically *towards oneself*. Teachers want their students for example to reflect upon their behaviour during a group work and the quality of the work they did. In four out of ten cases they want students to reflect on the value of information *(Critical thinking towards information)*. *Being able to handle uncertainty* is not an aim, while *Being conscious of the perspective of others* is in 15 percent of the modules intended.

Table 5-4

The Proportion of RR-learning Outcomes by Module Type (N = 31)

RR-learning outcomes Detailed RR-learning outcomes	Total	Module types **Facts only**	**Facts plus**	**Facts research-like illustrated**	**RB-facts explained & illustrated**	**Research process explained and applied**	**Insightful practice**
Results	1.00						
Underpinnings	0.35	0.00	0.20	1.00	1.00	0.40	0.22
Practical research skills	1.00						
Formulating a research question	0.03	0.00	0.20	0.00	0.00	0.00	0.00
Finding literature	0.03	0.00	0.00	0.00	0.00	0.00	0.11
Developing a design	0.13	0.00	0.00	0.00	0.33	0.60	0.00
Collecting data	0.26	0.00	0.00	0.67	0.00	0.60	0.33
Analysing data	0.10	0.00	0.00	0.67	0.00	0.00	0.11
Formulating a conclusion	0.32	0.00	0.00	0.67	0.00	0.00	0.89
Report	0.35	0.00	0.60	0.67	0.00	0.00	0.67
Research attitude	0.16	0.33	0.00	0.67	0.00	0.00	0.11
Systematic way of reasoning	0.68	1.00	1.00	0.33	1.00	0.40	0.44
Unspecified	0.13						
Competence to be a researcher	0.00						
Critical thinking	0.65	0.50	0.20	0.33	1.00	0.80	0.89
Critical thinking towards oneself	0.75	1.00	1.00	1.00	0.33	0.50	0.88
Critical thinking towards information	0.40	0.33	1.00	0.00	1.00	0.00	0.38
Conscious of perspective of others	0.15	0.33	0.00	0.00	0.00	0.00	0.25
Being able to handle uncertainty	0.00						
Unspecified	0.20						
Curiosity	0.39	0.33	0.20	0.33	0.67	0.60	0.33
Practice	0.00						
N	31	6	5	3	3	5	9

Note. The proportion of the detailed RR-learning outcomes, is based on the number of modules aiming at the overall RR-learning outcome. More than one interpretation per occurrence is possible. RR-learning outcomes with a proportion of 0 or 1 are not included in the comparison, neither are the unspecified detailed RR-outcomes.

5.3.2 Relation between the Module Types and RR-learning Outcomes

A comparison between the six module types and the RR-learning outcomes was made for *Underpinnings, Critical thinking*, *Curiosity,* and the detailed RR-learning outcomes of *Practical research skills* (Table 5-3). For the others such a comparison would be redundant.

In **Facts only** *Critical Thinking* is intended in half of the modules, *Curiosity* in two modules and *Underpinnings* in none. When *Critical thinking* is aimed at, it is predominantly *Towards oneself*. The intended practical research skill is mainly a S*ystematic way of reasoning*.

In **Facts plus** there is – apart for *Results*- only an outspoken attention for a S*ystematic way of reasoning* and *Report.*

The attention for *Underpinnings* in combination with limited attention for *Curiosity* and *Critical thinking* sets module type **Facts research-like illustrated** apart. Here a wide range of *Practical research skills* are intended.

Also all modules in **Research-based facts explained and illustrated** aim at *Underpinnings*, but here it is in combination with *Critical thinking (towards information)* and *Curiosity*. The intended *Practical Research Skill* is - as in **Facts only** and **Facts plus** - a *Systematic way of reasoning.*

In **Research process explained and applied**, *Critical thinking* is also important together with *Curiosity*. *Underpinnings* is intended but not as frequent as in the previous module type. Concerning the practical research skills, teachers aim most frequently at *Developing a design* and *Collecting data.*

In almost all modules of **Insightful practice**, *Critical thinking* is aimed, and here it is understood as *Critical thinking towards oneself*. *Underpinnings* or *Curiosity* are in some modules aimed at. The intended *Practical research skills* are most frequently *Formulating a conclusion* and *Report*.

5.4 Discussion

The aim of the study was to shed light on the relations between RI-approaches and RR-learning outcomes. The study revealed, as expected, that within a specific module type simultaneously different RR-learning outcomes are aimed at to a different extent. Notwithstanding this finding it is also the case that module types have "privileged" relationships with RR-learning outcomes. This is especially the case for the RR-learning outcomes *Underpinnings, Critical thinking* and *Curiosity*. In addition, a more fine-grained analysis disclosed that although all modules aim at the development of *Practical research skills*, there are substantial differences in the precise research skills aimed at. Similarly, the underlying interpretation of *Critical thinking* differed along the module types. For the remaining RR-learning outcomes such privileged links are not retrieved: in all modules *Results* are aimed at and in none the *Competence to be a researcher* or *Practice*.

Chapter 5

Interpreting these differences and similarities is not straightforward. The finding that all module types - also **Facts only** and **Facts plus-** aimed at the development of *Practical research skills*, is remarkable. The more detailed analysis showed however, that the interpretation of *Practical research skills* differed across module types. While in **Facts only**, **Facts plus** and **Research-based facts explained and illustrated**, it was predominantly interpreted as a S*ystematic way of reasoning*, the interpretation in the other module types was more diverse and more closely linked to research steps. For *Critical thinking* it was mainly in **Research-based facts explained and illustrated** that *Critical thinking* was oriented *towards information*. In the other modules *Critical thinking* mainly concerns *Critical thinking towards oneself*.

While there are differences in the intended RR-learning outcomes between **Facts research-like illustrated**, **Research-based facts explained and**

illustrated, **Research process explained and applied** and **Insightful practice**, none of the module types used one unique RI-approach, not used in another module type. In all these module types students were actively involved in research(-like) activities relevant for student learning (Segments- outcomes relevant). In addition, the presence of more RI-approaches does not necessarily imply that more RR-learning outcomes are aimed at. Although **Insightful practice** used a mix of five RI-approaches its prominent RR-learning outcomes were only *Critical thinking (towards oneself*) and the skill to write a C*onclusion*, while in other module types, there was a more outspoken attention for *Underpinnings* or *Curiosity*.

The difference between **Facts research-like illustrated** on the one hand and **Research process explained and applied** concerning the attention of *Underpinnings* was surprising. While they only differed with respect to one approach (Scientific facts and Research methods respectively*),* the attention for *Underpinnings* was higher in the former than in the latter. Nevertheless, one could assume that discussing research methods could also be a way to gain insight in the theoretical and methodological underpinnings of research (Lopatto, 2009).

The comparison between the RR-learning outcomes of **Facts research-like illustrated** and **Research-based facts explained and illustrated** is intriguing as well. While they both aimed at *Underpinnings*, the former aimed more at research skills such as *Collect data*, *Analyse*, *Conclusions* and *Report* while the attention for *Critical thinking* and a *Systematic way of reasoning* was more typical for the latter.

The findings of the present study show resemblances and differences with Visser-Wijnveen et al. (2010, 2012). **Facts only** and **Facts plus** might both be similar to type A, certainly if the assignments in type A do not show much resembles with research activities (see Table 5-1). The attention for *Critical* *thinking* in **Fact only** renders it different from type A. **Research-based facts explained and illustrated** shows parallels with type B because both focus on a critical attitude through the confrontation with research results and research-like activities. However, *Underpinnings* was not an intended learning outcome in type B (although Visser-Wijnveen et al. studied this learning outcome). **Insightful practice** resembles type C. In this study modules in which students participate in the research of the teacher (Type E) or follow in her footsteps (Type D) were absent. This absence may relate to the fact that here only first year modules were included. In the study of Visser-Wijnveen et al. type D and E were all more advanced modules. This supposition is supported by Van der Rijst and Jacobi (2010) who found authentic research activities within bachelor's programmes to be scarce. In Visser-Wijnveens study no module types like **Facts research-like illustrated** and **Research process explained and applied** were retrieved.

The study points out that a fine-grained analysis of research integration practices adds to the understanding of the complexities involved. The detailed distinction between RI-approaches made it possible to identify module types, which appear to be linked to specific combinations of RR-learning outcomes. Differences in frequency of RR-learning outcomes indicate the importance of details. Without the distinction between Facts, Scientific facts and Research-based facts less module types would have been identified, with as consequence less clear relations between the module types and the RR-learning outcomes. Similarly, the results of this study plead for the inclusion of fine-grained RR-learning outcomes in descriptions of research integration practices. Distinctive differences in what teachers aim at are more likely to be found at a fine-grained level than at a more general level, as shown in the differences between detailed interpretations of *Practical research skills* and *Critical thinking*. The findings illustrate the necessity of avoiding vague or general phrasing in favour of very precise use of language when discussing or studying research integration because otherwise essential differences will be missed (Trowler & Wareham, 2008; Spronken-Smith et al., 2012).

Research on research integration needs a detailed and manageable framework for describing accurately different research integration practices and for studying effects of differences in research integration. The RR-learning outcomes and the RI-approaches used in this study might contribute to the establishment of such a research framework.

Moreover, the results also indicate the value of looking at module types and not simply at individual RI-approaches. The analysis shows that although module types share the same RI-approach, they do not necessarily share the same RR-learning outcomes. Therefore it appears that the combination of the RI-approaches is a more functional unit of analysis to study research integration than individual RI-approaches. The description of educational practices at module level will probably be more informative than a description at the level of single approaches.

A practical implication of the study is the warning that a multitude of RI-approaches within one module is not necessarily desirable. The use of more RI-approaches does not imply the pursuit of more RR-learning outcomes. If the pursuit of the RR-learning outcomes is desired, this finding pleads for well-thought-out learning environments with careful considerations about the fit between the RI-approaches and the RR-learning outcomes. A lean learning environment might help to simultaneously aim at more learning outcomes.

While intriguing results are revealed, some caution is indicated. At least three limitations should be taken into account. One, the study is based on interviews. What teachers say they do and aim at is taken for granted. However, teachers may have difficulties to talk in a precise way about their teaching (Mälkki & Lindblom-Ylänne, 2012). Second, learning effects are not assessed. RR-learning outcomes solely refer to what teachers (tell they) aim at. There is no guarantee that students actual acquire the intended learning outcomes (Anderson & Krathwohl, 2001). It might also be that students learn more than teachers intended to. Students in the study of Visser-Wijnveen et al. (2012) reported to have learned more RR-learning outcomes than their teachers claim to aim at. It would therefore be valuable to study the relation between teachers RR-learning outcomes, teachers actions and students actual learning of RR-learning outcomes. Studies like the one of Van der Rijst (2009) on the association between teachers' intentions of research integration and student perceptions of it might be inspiring, as a step in that direction. Third the study is based on a limited set of only first year modules. This makes the established differences more easily prone to coincidence. The wider prevalence of the module types remains to be investigated. Studies of Visser-Wijnveen et al. (2010, 2012) and Zimbardi and Myatt (2012) indicate that module types where the approaches Segments-outcomes functional or Full study- outcomes functional are used, might be missing. Further research could investigate if the module types and the associate learning outcomes can also be retrieved in a larger and broader setting and indicate what module types are missing.

Chapter 6: Measuring Critical Thinking[9]

Although critical thinking (CT) is generally acknowledged as an important aim of higher education, no validated instrument to assess CT in Dutch is available. Moreover, most instruments are validated on a broad sample with people of diverse educational backgrounds. This possibly hampers the reliability of assessing effects of instructional interventions within educational programmes, where diversity is less. This study investigates the psychometric quality of a translation of the Cornell Critical Thinking Test (CCTT) and the Halpern Critical Thinking Assessment (HCTA) in a sample of Dutch-speaking freshmen majoring in educational sciences. Results show a higher content validity and preference by students for the HCTA. The CCTT, however, takes less time to administer and score, which makes it easier to use the CCTT on a larger scale. Neither of the two tests shows a high overall reliability. The strength of the correlations between the constructed-response items and the forced-choice items of the HCTA with the CCTT calls for further research on the precise relation between CT skills and dispositions and the ability of the HCTA to assess both independently.

[9] This chapter is submitted for publication. [Verburgh, François, Elen & Janssen (2013). *The assessment of critical thinking critically assessed in higher education: a validation study of the CCTT and the HCTA*. Manuscript under review].

6.1 Introduction

The development of critical thinking (CT) is generally acknowledged as an important aim of higher education (Bok, 2006; Moore, 2013; Vandermensbrugghe, 2004; Wood, Kitchener, & Jensen, 2002). Higher education graduates should be able to make decisions based on a well-thought consideration of available information. Research shows that students grow in their CT abilities during college (Astin, 1993; Gellin, 2003; Giancarlo & Facione, 2001; Miller, 1992), be it slowly and only moderately (Arum & Roksa, 2011; Pascarella & Terenzini, 2005; Pascarella et al., 2011; Wood & Kardash, 2002). There is however a lack of validated tests for CT development in Dutch speaking university students. The goal of the present study is therefore twofold: To investigate the psychometric properties of two commonly-used tests for CT in Flemish university students within one discipline and to assess their progress in CT using these two tests during one academic year. The results of the study are also valuable outside the Dutch language community because the study adds to the overall understanding of CT and its assessment difficulties. Moreover, the study is confined to students in one discipline in order to know the reliability of the instruments within more restricted populations. There is a demand of CT measures that are able to evaluate instructional interventions (Clifford, Boufal, & Krutz, 2004). Such instructional interventions are mostly conducted within one discipline and hence instruments need to be reliable within a restricted population.

In the following, the concept of CT is described first, afterwards current tests on CT are discussed. Finally, the purpose of the present study and its design are presented.

6.1.1 **The Concept of CT**

Despite the widespread agreement on the importance of the development of CT in students, agreement on its precise meaning is lacking (Abrami et al., 2008). The latter is exemplified by the variety of existing definitions on CT (Bailin et al., 1999; Butler, 2012; Erwin, 2000; Facione, 2010). At least two different considerations of the conceptualisation of CT can be discerned: (1) considering CT as discipline-specific and/or discipline-general and (2) considering CT as a set of skills or as a combination of skills with a disposition to be a "critical thinker".

Concerning the first aspect Moore (2004) distinguishes between two opposed movements in CT: the generalist movement and the discipline-specific movement. For the generalist movement, with Ennis (1989) as leading figure, CT is a set of cognitive abilities that can be taught independent from a specific content. The discipline-specific movement, with McPeck (1990) as leading figure, considers CT to be dependent on the problem area or the discipline under consideration. He argues that what counts as an appropriate use of scepticism in one discipline or context might be inappropriate in another. However, during the last decade the discussion between the two movements has become less prominent as most researchers agree that there are some general CT skills, which are applicable in various contexts, while familiarity with a discipline plays an important role too (Angeli & Valanides, 2009).

A second facet on which scholars differ in their conceptualisation of CT concerns the question whether CT is a set of skills or also a disposition (Ku, 2009). CT skills refer to, among others, rules of formal logic, consideration of multiple perspectives, induction, and deduction (Angeli & Valanides, 2009; Ku, 2009). In a dispositional viewpoint the motivation and the commitment of a person is included too, to see whether a person is able to recognise and willing to use the needed CT (Halpern, 2003). The dispositional viewpoint encompasses a more holistic view on CT, which stipulates that skills as well as other dispositional components together influence a person's CT performance (Ku, 2009).

6.1.2 **Tests of CT**

The diversity of conceptualisations of CT is mirrored in a diversity of available discipline-specific and discipline-general tests of CT (Saiz & Rivas, 2008). Even within discipline-general instruments, test developers depart from different conceptualisations of CT or place a different emphasis on particular aspects of CT. To give a few examples, the Reasoning about Current Issues Test (RCI) (King, Kitchener, & Wood, 2000) is based on the Reflective Judgement Model by King and Kitchener (2002). The Watson-Glaser Critical Thinking Appraisal (Watson & Glaser, 1980) aims at measuring CT as it is defined by Glaser (1941):

> (1) an attitude of being disposed to consider in a thoughtful way the problems and subjects that come within the range of one's experience; (2) knowledge of the methods of logical inquiry and reasoning; and (3) some skill in applying those methods. Critical thinking calls for a persistent effort to examine any belief or supposed form of knowledge in the light of the evidence that supports it and the further conclusions to which it tends. (p. 5)

The Cornell Critical Thinking Test (CCTT) (Ennis et al., 2005) is inspired by the Cornell/ Illinois model of CT. A fourth example is the Halpern Critical Thinking Assessment (HCTA) (Halpern, 2007; 2012), which is based on Halpern's definition (2003). As a final example, the California Critical Thinking Disposition Inventory (CCTDI) (Facione & Facione, 1992) claims to measure the inclination or disposition towards CT, as defined by Facione (1990).

In addition to the diversity in conceptualisations, a wide range of item formats is used in tests of CT. Commonly used instruments such as the CCTT (Ennis et al., 2005) use a forced-choice question format. However, in recent literature (Yeh, 2001; Ku, 2009; Saiz & Rivas, 2008) it is argued that a combination of forced-choice and constructed-response items is more suitable to measure CT because the constructed-response format allows to better grasp the dispositional aspect of CT. Unlike forced-choice questions, constructed-response questions enable to infer the respondent's reasoning behind an answer. Furthermore, the use of forced-choice questions may only indicate whether a respondent can recognise a correct answer or not, but it does not contain information about spontaneous answers from that respondent. As Ku (2009) argues, if a test is intended to measure dispositions as well as skills, the test ought to allow respondents to think spontaneously. The HCTA (Halpern, 2007; 2012) is a test that combines constructed-response and forced-choice items. Apart from the above mentioned item formats, still other formats have been used, such as interviews (e.g., the Reflective Judgement Interview; King & Kitchener, 1994), essays (e.g., Ennis-Weir critical thinking essay test; Ennis & Weir, 1985), a combination of essays and multiple-choice questions (e.g., the Critical Thinking Assessment Battery; American College Testing Program, 1991), and Likert-type statements (e.g., the Problem Solving Inventory; Heppner, 1988).

6.1.3 The Present Study

A validated instrument for assessing CT in Dutch-language students in higher education is lacking. Merely translating would not be sufficient to guarantee a valid instrument as cross-cultural assessment of generic skills as CT appears to be difficult (OECD, 2013). Moreover, most tests are validated on a broad population while most CT interventions are focused on students of one programme, and instruments need to be valid for a population with less variability. Therefore, the present study investigates the psychometric qualities of two instruments for assessing CT in students in higher education in Flanders, which is the Dutch-speaking part of Belgium. Two internationally used CT tests were selected and administered to a sample of freshmen (first-year students), majoring in educational sciences.

Three criteria were used for selecting the two instruments (Erwin, 2000; Cook et al., 1996). Firstly, the selected instrument had to measure CT ability irrespective of discipline-specific knowledge of students. Therefore, only discipline-general tests were considered.

Secondly, the underlying conception of CT of the selected instrument needed to fit with how CT is understood and taught in the field under consideration: higher education (HE) in Flanders. In accordance with Cook et al. (1996), a definition was established in close cooperation with representatives of higher education institutions in Flanders. This definition fits with how CT is understood and presumably taught in higher education in Flanders. The representatives agreed with the following shortened version of the definition of CT by Facione (2010), which considers CT as a combination of skills, leading a judgement, and dispositions, described as characteristics of the critical thinker:

> We understand critical thinking to be purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based.… The ideal critical thinker is habitually inquisitive, well-informed, trustful of reason, open-minded, flexible, fair-minded in evaluation, honest in facing personal biases, prudent in making judgments, willing to reconsider, clear about issues, orderly in complex matters, and diligent in seeking relevant information. (p. 2)

Finally, evidence of the psychometric quality of the instrument needed to be available.

According to these criteria, two Anglo-Saxon instruments were selected: the Cornell Critical Thinking Test-Level Z (CCTT) (Ennis, et al., 2005) and the Halpern Critical Thinking Assessment (HCTA) (Halpern, 2007; 2012). Although comparable in format, the CCTT was preferred above the Watson-Glaser Critical Thinking Appraisal (Watson & Glaser, 1980) because of a better match with the definition. The Reasoning about Current Issues Test (King, Kitchener, & Wood, 2000), although interesting in format, was not selected for this study because the interpretation of the scores was unclear.

The present study investigated the psychometric quality of both tests on four important aspects (Erwin, 2000; Cook et al., 1996): (1) reliability, (2) validity, (3) feasibility of the administration and scoring of the tests and (4) attractiveness of the test for the envisaged respondents.

Estimating the reliability of a test aims at estimating how much of variability in the scores of the test can be attributed to errors in the measurement and how much to true scores (Murphy & Davidshofer, 1991). There are different types of reliability. Here the interrater reliability and the internal consistency are investigated (Erwin, 2000). When a test is scored manually – as is the case for the HCTA – it is important that the scores are independent from the rater. Interrater reliability refers to this criterion (Neuman, 2011). Reliability also refers to the internal consistency of the test (Murphy & Davidshofer, 1991), which gives an indication whether the items that measure the same general construct produce similar scores.

The validity of a test refers to the extent to which the test accurately measures what it is supposed to measure (Kline, 1999; Murphy & Davidshofer, 1991; Vogt, 1999). There are different types of validity of which we will assess three: content validity, construct validity and criterion validity. Content validity concerns the degree to which the items in the test cover the domain of the construct being measured. Construct validity indicates the extent to which the variables of a test accurately measure the construct under consideration. It can be assessed at the level of the test by investigating the relation between the test scores and other tests, which is also called congruent validity. Construct validity can also be assessed at the level of the items of a test, by using factor analysis. Criterion validity can be defined as the ability of a test to make correct predictions. Therefore it is also often referred to as predictive validity. In our study, both tests were used to assess the progress in CT for freshmen between the beginning and the end of the academic year. In fact, these data can be considered as part of the investigation of the criterion validity of both tests.

Finally, the feasibility of the test concerns the ease of the test to administer and analyse. The attractiveness of a test relates to the extent to which respondents like the test (Cook et al., 1996). It is assumed that the more attractive respondents find the test, the more they will be willing to commit to taking the test.

6.2 Method

6.2.1 Instruments

Assessment of CT

HCTA

The HCTA (Halpern, 2007; 2012) is a recently developed discipline-general test which exists of 25 descriptions of daily-life situations. Each situation is offered twice to the respondents: A first time followed by an open-ended question, where students have to construct their own answer (constructed-response item) and a second time followed by a forced-choice question (forced-choice item). The forced-choice items have different formats: multiple-choice questions with one or with more than one correct answer; rating questions with a Likert-type or with a yes/no scale; and matching questions. The maximum score per item ranges from one to ten. The test aims at measuring five categories of CT. Each category is measured in five situations. However, when calculating a total test score, the contribution of each category differs: (1) Hypothesis testing (24%), (2) Verbal reasoning (12%), (3) Analysis of arguments (21%), (4) Use of likelihood and uncertainty (12%), and (5) Decision making and problem solving (31%).

The HCTA results in thirteen different scores. Apart from the total score, there is a total score for the constructed-response items (Constructed-response part) and the forced-choice items (Forced-choice part). In addition, there are five subscores in each category, both for the constructed-response items and the forced-choice items.

For the present study, the five categories of the HCTA show a good correspondence with the first part of the developed definition of CT. By using the constructed-response questions, the HCTA claims to be able to measure CT dispositions (Ku & Ho, 2010). Because the HCTA is a very recent instrument, data about the psychometric features are mainly limited to the test manual and to research in close cooperation with the author. According to the test manual, the HCTA has a high internal consistency (Cronbach's α = .88 for the total score) (Halpern, 2012). The Constructed-response part and the Forced-choice part separately show high internal consistencies as well (respectively α =.84 and α = .79). The respondents in the sample reported in the manual have diverse educational backgrounds and their age ranges from 18 to 72. Reliability analyses of translations into Chinese and Spanish found Cronbach's α ranging from .69 to .77 for the overall test and low reliabilities for the subscales (α = .34 to .64) (Chan, Ho, & Ku, 2011; Ku & Ho, 2010; Nieto, Saiz, & Orgaz, 2009). In these studies the sample consisted of students of different years and of different disciplines.

Correlations between the Constructed-response part and the Forced-choice part indicate that both parts measure related but distinct constructs (Halpern, 2012). Factor analyses point in the direction of a ten-factor structure (the five categories with a distinction between constructed-response and forced-choice items) (Halpern, 2012). Evidence for criterion validity has been established by Butler et al. (2012), who found that the HCTA predicts real-world outcomes of CT. In addition there are significant correlations between epistemological beliefs and the HCTA (Chan et al., 2011).

CCTT

The CCTT (Ennis et al., 2005) is a discipline-general test, intended for strong students in upper secondary education, students in higher education, and adults. Developed in 1985, it is a widely used instrument for assessing CT. It aims at measuring five aspects of CT: Deduction; Semantics; Observation and credibility of sources; Induction; Definition and assumption identification. Each aspect is measured with a separate section in the test, but Induction is split into two sections, namely on the use of Induction in hypothesis testing and in Planning experiments. The test contains 52 items, all of which are in a forced-choice format. Similar to the HCTA, the CCTT has a relatively good correspondence with the first part of the definition. On the other hand, the second part of the definition, which captures the concept of CT as a disposition, is lacking, because the instrument –as most other instruments- only uses multiple-choice questions.

Regarding its reliability, the CCTT's manual reports split-half reliabilities between $r = .49$ to .80 and Kuder-Richardson reliabilities between KR = .50 and .76 (Ennis et al., 2005). The respondents in the studies were mostly undergraduate students or graduate students, mostly within one discipline. Taking all studies together, the CCTT was evaluated in a broad range of different institutions of higher education. Erwin (2000) reports internal consistency values of $\alpha = .58$ in a sample of freshmen and of $\alpha = .72$ in a sample of sophomores.

The content validity of the CCTT was assessed by the members of the Illinois Critical Thinking Project, who agreed that the items of the CCTT measure CT as defined by the authors (Ennis et al., 2005). In addition, there are positive indications for criterion validity. For example, the correlation with the reflective judgement interview of King and Kitchener is .46 (King, Wood, & Mines, 1990).

Translation Procedures

Both tests (Ennis et al., 2005; Halpern, 2007; 2012) were translated into Dutch, following the International Test Commission guidelines for translating and adapting tests (International Test Commission, 2010). This translation process consisted of several steps, following Wang, Lee, and Fetzer (2006). After a first translation was made, a pilot study (N = 5) was conducted, in which respondents filled out a translated test and were asked about their comments on the items of the tests during cognitive interviews. Adaptations to the translations were made. Next, a validation of the translation was done using the translation-back technique (Maneesriwongul & Dixon, 2004). In this procedure the Dutch versions were back translated in English by a third person and the translated version was compared with the original English version. The differences in both versions were discussed and adaptations were made to the Dutch translations. Finally, both Dutch versions of the tests were administered to two different try-out groups of students in order to fine-tune the translation and cultural adaptation (N = 66 for the CCTT; N = 40 for the HCTA).

In order to establish cultural appropriateness for our population (Wang et al., 2006), several items of the HCTA were slightly changed during the translation process compared to the original English version. For example, one situation acts about a presidential candidate. This was changed into "politician", because the investigated population does not have a president. In another situation respondents have to estimate the chances of a young woman to become a famous actress in Hollywood. The Dutch translation specifies that the young woman is American, because the tryout showed that some students' answers were influenced by the assumption that the woman came from a small, non-English speaking country, and that this lowered her chances. Although this was a correct inference, this answer was not intended by the original test.

For the CCTT, one item (item 21), was removed from the translated version, because this item is mainly based on the double meaning of the word "drugs". In Dutch there is no equivalent with the same double meaning.

Assessment of Attractiveness

The attractiveness of the tests was measured with a self-developed questionnaire consisting of three parts. In the first part of the questionnaire, students were asked to evaluate each test separately on a seven-point response scale (ranging from totally disagree to totally agree) concerning its difficulty, attractiveness, time and amount of reading/writing necessary to fill in the test. For the CCTT the use of the forced-choice questions only was also evaluated. Next, students had to make a forced choice between the two instruments, regarding features as being interesting, difficult, and possibility to show thinking ability. Finally, students could freely comment on the tests and explain their preference.

6.2.2 **Procedure**

Both tests were administered twice as a compulsory part of a first year module. The tests were first administered at the beginning of the first semester (November) and again at the end of the second semester (May). The CCTT was administered on paper, the HCTA online. For each test there were different collective sessions, from which students could choose their most convenient moment. In November the collective sessions for the HCTA and the CCTT were mixed. In May all the CCTT sessions were planned before the HCTA-sessions. After finishing the HCTA in May, students were asked to fill in the attractiveness questionnaire. Each session started collectively. When students had finished, they could leave the room.

The answers to the CCTT and to the forced-choice questions of the HCTA were scored with the key of the manual. The answers on the constructed-response questions of the HCTA were scored according to the Vienna Test System (Halpern, 2012), accompanied with the manual with examples. The Vienna Test System guides the rater through the respondent's answers on the constructed-response questions, with a series of prompts to be answered with "yes" or "no" or "yes", "no", "somewhat". This system is intended to increase the speed and the reliability of the scoring. After the establishment of interrater reliability (see section interrater reliability for details), the questions were scored by one rater.

In between the two administration moments, a workshop on CT with twenty representatives of different higher education institutions was held. These representatives were partly different from the persons who developed the definition of CT. They were first asked to individually envisage a person in their own field who thinks critically and to write down what the person does. Then they had to compare in small groups the activities they wrote down and label the activities in abstract words. Afterwards, in a plenary session they compared the activities with the used definition and with both instruments.

Participants

The participants were freshmen majoring in educational sciences at the KU Leuven (mean age = 18.2), Flanders, Belgium. In total 179 students filled out at least one test, of which 154 filled out both tests twice. The majority of the respondents were female, which is a normal situation for educational sciences in Flanders (see Table 6-1).

Table 6-1

Number of Participants in Each Administration Period, by Gender

Test	Gender	November	May
HCTA	Female	162	149
	Male	10	10
	Total	172	159
CCTT	Female	163	152
	Male	11	11
	Total	174	163

6.2.3 **Analyses**

Reliability

For the constructed-response questions of the HCTA interrater reliability was investigated. Two raters individually scored the responses of 20 students. Afterwards, the differences were discussed in order to make sure that the prompts of the Vienna Testing System were understood in the same way. Next, the responses of 50 randomly selected students were scored by two raters. The results were compared and differences were discussed. For each item the proportion of equal scores and the weighted Cohens κ were calculated. It was decided beforehand that if questions had a proportion equal scores lower than .7, the responses of 50 additional students would be scored. During the establishment of interrater reliability the Dutch version of the scoring guide was elaborated with more examples to ease the scoring and make it more transparent.

In order to make comparisons with the interrater reliabilities reported in the manual (Halpern, 2012), the correlations between the subscale scores of the two raters were looked at. In addition, the effect of the rater on the means of the subscale scores was calculated by using a paired samples *t*-test.

The internal consistency was measured using the Cronbach's α. For both tests it was calculated separately for the November and May administration.

For the HCTA it was calculated for the overall result, for the constructed-response part, for the forced-choice part, for the five categories and also for the ten subscales. Because in the HCTA the maximum score differs per item, two types of Cronbach's α were calculated: the normal Cronbach's α and the Cronbach's α of the standardised items. In addition to the analysis based on the scores of the individual items, the internal consistency was also calculated using the sum of the items of each subscale as a variable, because this approach was followed in the manual of the HCTA (Halpern, 2012). This additional calculation allows to compare our results with the results in the manual.

For the CCTT, Cronbach's α of section IV and V and of section VI and VII were calculated for both sections together, because these sections each measure the same or highly comparable aspects of the CCTT (respectively: Induction and Assumption identification).

Validity

In order to assess the content validity it was evaluated how every single aspect of the developed definition on CT was covered within both tests. These aspects of the developed definition are (1) purposeful, self-regulatory judgement, (2) interpretation, (3) analysis, (4) evaluation and (5) inference, (6) explanation of evidential and conceptual considerations, (7) explanation of methodological considerations, (8) explanation of criteriological considerations (9) explanation of contextual considerations, and (10) the "ideal critical thinker". Additional information on a close match between the conceptualisation of CT in Flanders and the instruments was gained during the workshop on CT with twenty representatives of different higher education institutions.

In the present study, the correlation between both tests in the same administration period can be used to assess construct validity. In addition to the observed correlation, also the correlation with correction for attenuation is used (Muchinsky, 1996). This correction allows to correct for a lack of perfect reliability, due to measurement errors which are inherent to empirical measures. Due to these measurement errors, the observed correlation is lower than the true correlation (Scheerens, Glas, & Thomas, 2003). For the HCTA the total score is taken into consideration, as well as the constructed-response part and the forced-choice part separately. It is expected that the CCTT will correlate higher with the total score and the forced-choice part of the HCTA than with the constructed-response part, because the constructed-response part is intended to measure also the dispositional aspect of CT, whereas the others are more restricted to CT skills.

In addition, the correlation between the constructed-response and the forced-choice part of the HCTA during the same administration period was looked at. Again, both the observed correlation and the correlation with correction for attenuation were considered. Because the two parts both measure aspects of CT, but with a different focus, a moderately strong correlation is expected.

In order to assess construct validity of the HCTA at item level a principal component analysis (PCA) with an oblique rotation was planned on the 50 items for both the November and the May data. Based on the logic of the test either five or two interdependent factors are expected (reflecting the skills and disposition measured in the respectively forced-choices items and the constructed-response items). Before the analysis the Kaiser-Meyer-Olkin (KMO) measure was used to verify the sample adequacy for a PCA. The latter is confirmed when KMO > .5 (Field, 2009).

In order to assess the dimensionality of the CCTT, a Multidimensional Item Response Theory (MIRT) model was used (e.g., Bock, Gibbons, & Muraki, 1988). A MIRT model is also called "item factor analysis". It is similar to a classical factor analysis in that it tries to assess the underlying dimensionality of a test. However, a MIRT model models the data set of the person by item responses directly, whereas a classical factor analysis models the correlations over persons between the responses on the items of a test. Given this different approach, MIRT models are more apt to derive the dimensionality of a test with dichotomous items, because for such items, the correlation matrix is more difficult to assess (Bock et al., 1988). The MIRT model was estimated using the R package mirt (Chalmers, 2012). Models with one to five dimensions were compared using the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) (Burnham & Anderson, 2002; 2004). The preference of one model above the other depends on the distance of the model with the data. The smaller the distance (the smaller the value of the criterion), the better the fit between the model and the data (Linhart & Zucchini, 1986). The best fit is expected for a model with five dimensions, given the five aspects of CT underlying the CCTT.

Criterion validity was assessed by looking at the correlation between the scores in November and May on the same test and by calculating the progress of individual students across both assessments.

Feasibility

The time to administer and to score the test were considered as criteria to assess the feasibility.

Attractiveness

With paired samples *t*-tests the appreciation of both tests was compared. In addition, the proportion of students preferring one test above the other was considered.

6.3 Results

6.3.1 Descriptive Statistics

Table 6-2 and 6-3 describe the CT performance in November and May on the HCTA and the CCTT respectively. On average, the total score was 116.08 in November and 120.32 in May on the HCTA. The difficulty level of the items varied. Some items were very easy (e.g., in May almost all students answered the forced-choice question of Situation 9 correctly). Other items were difficult (e.g., in November almost 9 out of 10 students scored no points on the forced-choice question of Situation 17). Similarly the subscales differed in difficulty: Items of the Argument analysis forced-choice subscale seemed easier for students than items of the Hypothesis testing constructed-response subscale. The average difficulty varied between .15 and .99 proportion correct answers.

The average score on the CCTT was 27.13 in November and 27.53 in May. The difficulty levels of the items differed. Some items were hardly answered correctly (proportion correct answers of .07) while others were almost always answered correctly (proportion correct answers of .96). The difficulty of the subscales also differed: Students answered more items correctly on the Deduction scale than they did on the Meaning and fallacies scale.

Table 6-2

Descriptive Statistics for the HCTA (N = 155)

Scale	Max	November				May						
Subscale	value	*M*	*SD*	Min	Max	*M*	*SD*	Min	Max	*t*	*df*	*p*
Total	195	116.08	10.11	84	140	120.32	10.88	92	147	5.144	154	.000
Constructed response items (C)	95	49.99	6.68	31	66	52.81	7.40	37	68	4.885	154	.000
Forced-choice items (R)	99	66.09	5.33	52	77	67.52	5.73	53	83	3.058	154	.003
Hypothesis testing	45	26.35	4.14	16	38	27.43	3.83	17	36	2.990	154	.003
Hypothesis testing-C	18	9.49	2.42	3	16	10.54	2.44	3	16	4.607	154	.000
Hypothesis testing-R	27	16.86	2.75	9	23	16.90	2.66	11	23	0.130	154	.897
Verbal reasoning	22	11.26	2.49	5	18	11.52	2.53	4	18	1.177	154	.241
Verbal reasoning-C	15	7.06	2.32	2	14	7.20	2.25	2	13	0.668	154	.505
Verbal reasoning-R	7	4.20	0.86	2	6	4.32	0.94	1	7	1.364	154	.174
Argument analysis	42	25.10	4.45	14	35	26.88	4.30	13	36	4.360	154	.000
Argument analysis-C	23	12.10	2.94	5	19	12.81	3.17	4	20	-0.621	154	.535
Argument analysis-R	19	12.99	2.42	6	18	14.08	2.20	7	19	4.609	154	.000
Likelihood and uncertainty	24	13.96	2.99	4	21	14.75	3.10	6	21	3.115	154	.002
Likelihood and uncertainty-C	17	9.30	2.50	1	15	9.88	2.77	2	15	2.488	154	.014
Likelihood and uncertainty-R	7	4.66	1.08	2	7	4.87	0.90	2	7	2.237	154	.027
Decision making and problem solving skills	61	39.41	4.18	26	49	39.74	4.27	30	51	0.838	154	.403
Decision making and problem solving skills-C	22	12.04	2.81	6	18	12.39	2.70	6	20	1.295	154	.197
Decision making and problem solving skills-R	39	27.37	2.54	20	33	27.35	3.01	20	34	-0.072	154	.943

Table 6-3

Descriptive Statistics for the CCTT (N = 157)

Scale	Max value	November *M*	*SD*	Min	Max	May *M*	*SD*	Min	Max	*t*	*df*	*p*
Total	51	27.13	4.24	15	37	27.53	4.39	16	39	1.049	156	.296
I Deduction	10	6.26	1.37	3	10	6.11	1.51	2	9	-0.947	156	.345
II Meaning & fallacies	10	3.45	1.49	0	8	3.95	1.49	1	8	3.419	156	.001
III Observation & credibility of sources	4	2.31	1.00	0	4	2.38	1.04	0	4	0.710	156	.479
IV/V Induction	17	9.50	1.91	5	14	9.32	1.83	4	14	-1.068	156	.287
VI/VII Definition & assumption identification	10	5.60	1.64	1	9	5.76	1.69	1	9	1.058	156	.292

6.3.2 **Reliability**

Interrater Reliability

At item level, there was a high proportion of equal scores and satisfying weighted Cohen's κ, except for four situations (Situations 3, 6, 15 and 25), as can be seen in Table 6-4. For these four items the responses of 50 additional students were scored and then the results were satisfactory (indicated between brackets).

Table 6-5 shows the correlations between the scores of both raters on the five subscales. According to Cohen (1988) these correlations were large. They were comparable or larger than those reported in the manual. However, the paired sample *t*-tests revealed – in contrast with the manual – a significant effect of rater for the constructed-response part, with a small effect size. For the subscales Hypothesis testing and Likelihood and uncertainty there was also a significant effect with a medium effect size. The scatter plots revealed that on these scales one rater systematically scored somewhat higher than the other.

Table 6-4

Interrater Reliabilities for the Constructed-Response Items of the HCTA (N = 50)

Situation	Proportion equal scores	Weighted Cohen's kappa
Situation 1	.86	.77
Situation 2	.70	.50
Situation 3	.58 (.72)*	(.77)
Situation 4	.88	.85
Situation 5	.72	.62
Situation 6	.56 (.76)	(.72)
Situation 7	.84	.78
Situation 8	.86	.77
Situation 9	.86	.76
Situation 10	.94	.83
Situation 11	.90	.89
Situation 12	.86	.90
Situation 13	.70	.58
Situation 14	.92	.90
Situation 15	.66 (.68)	(.62)
Situation 16	.78	.72
Situation 17	.88	/
Situation 18	.74	.74
Situation 19	.72	.75
Situation 20	.88	.75
Situation 21	.74	.73
Situation 22	.80	.71
Situation 23	.70	.68
Situation 24	.98	.94
Situation 25	.66 (.74)	(.65)

Note. For question 17 it was impossible to calculate the weighted Cohen's kappa because not all values were present. *The results between brackets are the results of a second set of 50 responses.

Table 6-5

Interrater Correlations and Differences between the Means between Two Raters on the Constructed-Response Subscales of the HCTA

	Interrater Correlations		*Differences between the means of two raters*			
Subscale –constructed-response items	*r* in the sample	*r* in the manual	*t*	*df*	*p*	Cohen's *d*
Thinking as hypothesis testing	.85	.75	4.632	49	.000	.37
Verbal reasoning	.84	.60	-0.414	49	.681	-.03
Argument analysis	.88	.70	-1.531	49	.132	-.10
Likelihood and uncertainty	.89	.82	3.357	49	.002	.23
Decision making and problem solving	.84	.53	1.014	49	.315	.08
Constructed-response part	.93	.83	3.063	49	.004	.16

Internal Consistency

Table 6-6

Internal Consistencies (Cronbach's α)

	November		May	
Scale Subscale	α	Standardised α	α	Standardised α
HCTA				
Total	.53	.55	.64	.64
Constructed-response part (C)	.34	.37	.53	.53
Forced-choice part (F)	.35	.37	.48	.44
Hypothesis testing	.42	.46	.39	.44
Verbal reasoning	.21	.17	.28	.28
Argument analysis	.37	.38	.42	.44
Likelihood and uncertainty	.18	.31	.31	.34
Decision making and problem solving	.26	.25	.35	.32
Hypothesis testing-C	.21	.20	.27	.28
Hypothesis testing-F	.30	.34	.39	.39
Verbal reasoning-C	.24	.24	.25	.27
Verbal reasoning- F	.01	-.04	.20	.19
Argument analysis-C	.16	.17	.31	.30
Argument analysis-F	.10	.13	.17	.24
Likelihood and uncertainty-C	-.02	.03	.26	.31
Likelihood and uncertainty-F	.17	.17	-.12	.04
Decision making and problem solving-C	.19	.20	.26	.24
Decision making and problem solving-F	-.01	-.04	.24	.21
CCTT				
Total	.52		.52	
I Deduction	.07		.17	
II Meaning & fallacies	.30		.14	
III Observation & credibility of sources	.17		.25	
IV/V Induction	.22		.15	
VI/VII Definition & assumption identification	.32		.42	

Table 6-6 presents the internal consistency of the HCTA and the CCTT for the overall test and the different subscales for the two test administrations. There was not much difference between both types of Cronbach's α. Therefore, the values of the standard Cronbach's α are reported below.

The November administration of the HCTA had an alpha of .53, which is a moderate internal consistency (Cohen, 1988). The internal consistencies of the Constructed-response and Forced-choice part separately were low (respectively α = .34 and .35) as well as for the five categories of CT (α < .4).

The internal consistency of the May administration of the HCTA was .64, which is acceptable (Cohen, 1988). In contrast to the November data the Constructed-response part and the Forced-choice part had a moderate internal consistency (respectively α = .53 and α = .49). The internal consistencies of the separate CT categories of the HCTA were somewhat higher than in November, but still low to moderate (α < .43).

The internal consistencies of the subscales were also low in the two test administrations (α < .40). Because of the low number of items in the subscales, the inter-item correlations were considered, but these were also low (predominantly r < .20). For each scale there was a limited number of items which correlated sufficiently with the total scale (r > .25) (Ebel & Frisbie, 1991). In addition, for each scale there were items that correlate negatively –but close to zero - with the scale total. The items with a sufficient correlation and with negative correlations differed between the two administration moments. Items which were very easy or difficult generally had a low correlation with the test total.

When calculating the internal consistency similarly as in the manual (taking the scores of the subscales as items), the α's were still low, with the exception of the overall score in May (Table 6-7).

Table 6-7

Internal Consistency of the HCTA with Scales as Variable

	Sample		
Scale	November	May	Manual
Total	.50	.58	.88
Constructed-response part	.28	.46	.84
Forced-choice part	.32	.40	.79

For the CCTT, the internal consistency was moderate in November ($\alpha = .52$) (Table 6-6). Only two items correlated sufficiently with the total scale ($r > .25$) and one item correlated negatively with the total scale. The Cronbach's *α's* for the CCTT subscales were low ($\alpha < .30$). The internal consistency based on the May data was the same as for the November data ($\alpha = .52$), with five items sufficiently correlating with the total scale ($r > .25$) and seven items with a negative correlation with the total scale. Again, the α for the subscales were low ($\alpha < .30$), with the exception of Assumption identification, which was somewhat higher ($\alpha = .42$).

6.3.3 **Validity**

Content Validity

Table 6-8 shows the results of the comparison between both tests and the developed definition. Both tests adequately mirrored the part of the definition on CT skills. The second part, with the dispositional aspects of CT, was measured by the constructed-response items of the HCTA, where respondents had to formulate their own answers. This was not the case with the CCTT.

During the workshop with the representatives of different higher education institutions, a close match between the perception of CT in Flanders' higher education on the one hand and both tests on the other hand was found. All CT activities they identified were covered with the intended categories or sections of both tests.

Table 6-8

Content Validity: Match between the Used definition of CT and the Elements Measured in the Two Tests

Aspects of CT	HCTA	CCTT
Skills		
Purposeful, self-regulatory judgment	all	all
Interpretation,	Verbal reasoning	VI Identification of definitions and assumptions
		VII Identification of assumptions
Analysis,	Argument analysis	I Deduction
	+ situation 1 and 2	
Evaluation	Verbal reasoning	II Meaning and fallacies
Inference	Argument analysis	I Deduction
	+ situation 1 and 2	
Explanation of evidential and conceptual considerations	Verbal reasoning	VI Identification of definitions and assumptions
		VII Identification of assumptions
Explanation of methodological considerations	Hypothesis testing	III Observation and credibility of sources
	Likelihood and uncertainty	IV Induction (hypothesis testing)
		V Induction (planning experiments)
Explanation of criteriological considerations	Decision making and problem solving	II Meaning and fallacies
	Likelihood and uncertainty	
Explanation of contextual considerations	Decision making and problem solving	IV Induction (hypothesis testing)
	Likelihood and uncertainty	V Induction (planning experiments)
Disposition		
The ideal critical thinker	Constructed-response items	/

Construct Validity

Table 6-9 shows the correlations between both tests in November in the upper left corner and in May in the lower right corner. All correlations were significantly different from zero, with a small strength (Cohen, 1988). The correlations corrected for attenuation indicated a relationship of medium strength between both tests.

In November as well as in May the correlation between the Constructed-response and the Forced-choice parts of the HCTA was significantly different from zero, with a medium strength (Cohen, 1988). The correlation corrected for attenuation indicated a strong relation between both parts.

Table 6-9

Correlation between Both Tests in November and in May (with Correction for Attenuation)

	1	2	3	4	5	6	7	8
Nov								
1. CCTT								
2. HCTA total[1]	.25**							
	(.35)							
3. HCTA C	.21**							
	(.36)							
4. HCTA F	.22**		.41**					
	(.37)		(.70)					
May								
5. CCTT	.37**	.36**	.32**	.28**				
	(.51)	(.49)	(.44)	(.39)				
6. HCTA total	.18*	.52**	.43**	.46**	.34**			
	(.22)	(.66)	(.54)	(.57)	(.42)			
7. HCTA C	.18*	.49**	.49**	.32	.33**			
	(.25)	(.67)	(.67)	(.44)	(.45)			
8. HCTA F	.11	.36**	.18*	.45**	.22**		.36**	
	(.15)	(.52)	(.27)	(.65)	(.32)		(.52)	

* Significant at .05-level (2-tailed), ** Significant at .01-level (2-tailed); [1]The correlation of the total of the HCTA with the Constructed-response part and the forced-choice part is left out of the analysis because the total is partly composed of each part.

The Kaiser-Meyer-Olkin measure to verify the sample adequacy for a PCA indicated that the November sample was inadequate, KMO= .45 (Field, 2009). The Kaiser-Meyer-Olkin measure for the May sample was slightly better but still under the criterion of .5, KMO = .49. Therefore it was decided to skip the analyses.

For the CCTT, the fit statistics of the five exploratory factor solutions indicated that the unidimensional solution was the most parsimonious (Table 6-10). The AIC and BIC values were the lowest for the solution with one dimension, except for the AIC in November, where the two dimensions solution was slightly lower than the one dimension solution.

Table 6-10

Fit Statistics for MIRT Models with One to Five dimensions, for the November and the May Administration of the HCTA

	AIC		BIC	
Dimensions	November	May	November	May
1	9626	9113	9948	9429
2	9617	9113	10098	9583
3	9644	9139	10279	9761
4	9707	9175	10493	9945
5	9765	9249	10700	10164

Criterion Validity

The correlations between the November and May results on the same (part of the) test were significantly different from zero (lower part of Table 6-9). When the correlations were corrected for attenuation, there was an indication of a strong relationship between the two test moments.

For both tests, differences between both moments were investigated with paired sample *t*-tests. For the HCTA there was a significant growth. This growth was confined to specific subcategories. Students did not advance in Verbal reasoning and Decision making and problem solving skills. Neither did they advance on the

Hypothesis testing forced-choice subscale and on the Argument analysis constructed-response subscale. For the CCTT there was no growth.

6.3.4 **Feasibility**

The average testing time of the CCTT was 54 minutes. Students were rather neutral in their opinion when asked if the CCTT takes too long to complete (Table 6-11). About half of the students at least slightly agreed that the CCTT took too long. On average, the respondents needed about 80 minutes to complete the HCTA. When asked about their opinion on test administration time, on average students slightly agreed that it took too long to fill in the HCTA. The difference in opinion about test completion time between both instruments was significant ($p < .001$).

The scores of the CCTT and the forced-choice items of the HCTA could be calculated automatically because of the question format. The scoring of the constructed-response items of the HCTA could not be done automatically and was time consuming. It required some practice in order to score systematically. Estimated scoring time after a short training period was 15 min per test.

Table 6-11

Students' Opinions on the HCTA and the CCTT (N = 132)

	HCTA		CCTT				
Question	*M*	*SD*	*M*	*SD*	*t*	*df*	*p*
Test is difficult	4.84	1.13	5.52	.94	5,956	131	,000
Test is too difficult	3.20	1.39	4.11	1.56	5,472	130	,000
Test is fascinating	4.54	1.26	3.49	1.28	7,418	127	,000
Takes too long to fill in	5.05	1.50	4.41	1.55	3,832	131	,000
Too much reading to do my best	4.14	1.63	4.34	1.66	1,411	130	,161
Too much writing to do my best	3.12	1.55	/	/			
Would have done better in case of constructed-response opportunity	/	/	3.26	1.73			

6.3.5 **Attractiveness**

On average, students slightly agreed that both instruments were difficult, but they found the CCTT significantly more difficult (Table 6-11). Students slightly disagreed that the HCTA was too difficult, while agreeing more that the CCTT was too difficult. This difference was significant ($p < .001$). On average students slightly disagreed that the CCTT was fascinating, while being neutral about the HCTA, which was again a significant difference ($p < .001$). Finally, for both tests students were neutral about whether they could have done better on the test if less reading would have been involved. There was no difference between both instruments. They slightly disagreed that they could have done better on the HCTA if they had to write less. They slightly disagreed that they could have done better on the CCTT if they would have had the opportunity to construct an answer themselves instead of being forced to choose an answer from a list.

Students were also asked about their test preference (Table 6-12). About three-quarters of the respondents preferred the HCTA above the CCTT ($p < .001$). The HCTA was found to be the most interesting, fascinating and motivating test (p < .001). To a lesser degree (p < .01), the HCTA was also deemed to be the most challenging test and the test with the greatest possibilities to demonstrate their thinking ability. Most students found the CCTT the most difficult test (p < .001). According to our respondents, the HCTA owed its appeal to its use of more familiar, recognisable everyday situations and constructed-response items. These items gave students the opportunity to express their own opinion, what they highly appreciated. Opposed to the HCTA, the CCTT's situations were not familiar and its questions were regarded to be abstract. Students frequently mentioned the time needed to complete the HCTA and the fact that they had to write a lot themselves as negative points of this test. Hence it comes as no surprise that the nature of the multiple response format of the CCTT was cited most often as its major advance.

Table 6-12

Students' Preference of One of the Instruments (%)(N = 132)

Choice	CCTT	HCTA
Most interesting	22.2	77.8
Most difficult	83.9	16.1
Most fascinating	23.5	76.5
Most challenging	37.0	63.0
Most motivating (first time)	27.8	72.2
Most motivating (second time)	27.2	72.8
Best showing my thinking ability	39.2	60.8
Preferred test	25.5	74.5

6.4 Discussion

This study compared two internationally widely used instruments to measure CT on their reliability, validity, feasibility and attractiveness for students in higher education: the HCTA and the CCTT.

6.4.1 Reliability

The interrater reliability established in the study was satisfying, as measured in the high proportions equal scores and good to very good weighted Cohen's κ's, for almost all items (Jakobsson & Westergren, 2005). In addition, the correlations between subscale totals of the two raters were strong. They were higher or comparable to those reported in the test manual (Halpern, 2012). However, in contrast to the manual, there was a significant effect of the rater on the scores. This might be caused by the fact that one rater scored systematically a little higher than the other. Despite this effect of the rater, the interrater reliability was satisfying. It indicates objectivity in the test scores.

The internal consistencies found in this study were low. This was not affected by the different weights of the items for calculating the total score, as indicated with the highly similar internal consistency on the raw or standardised item responses. Even when taking into account that the concept of CT is rather complex, the reported reliability measures are insufficient. The internal consistencies on the tests found in the test manuals and in literature could not be replicated for the Dutch translations of the tests in the investigated research group. The difference between the Cronbach's α in this study and other studies might be due to the effect of the restriction of range or differences in the breadth of the population under investigation (Scheerens et al., 2003). Reliability is dependent on the variance: the higher the variance, the higher the reliability. When a diverse population is considered, the variance is higher than in a more restricted population, which has a positive effect on the reliability (Scheerens et al., 2003). The reliability of the HCTA, reported in the manual, was assessed in a sample of respondents of different age groups and with diverse educational backgrounds (Halpern, 2012), while the investigated population in this study is restricted to freshmen of one major at one university. A comparison between the descriptive statistics of the standardisation sample in the manual and of our sample indicated that although the means were comparable, the variances in the manual were indeed larger than in the present study, supporting the restriction of range explanation. Furthermore most other studies describing reliabilities of the HCTA and the CCTT used broader samples than freshmen in one discipline.

6.4.2 **Validity**

The content validity of both tests is sufficient. Although neither of the two instruments captures the whole concept, they both match closely to the definition used by higher education staff members in Flanders. The HCTA has an extra strength by using a combination of constructed-response questions and the forced-choice questions, because the latter makes it possible to measure the dispositional aspects of CT.

Concerning the construct validity, the correlations between both tests during the same assessment period indicated a medium strong relationship. When the correlations are corrected for attenuation, the relationships indicated a medium to strong relationship. This suggests that the CCTT and the HCTA are –at least partially- measuring the same constructs and that the lack of correlation is partly due to the lack of reliability. Although the HCTA had a higher content validity because of its intention to measure the dispositional component of CT, this difference was not reflected in differences in strength of the correlations. In May the relationship with the CCTT was even stronger for the Constructed-response part than for the Forced-choice part, while theoretically the opposite could be expected. The same holds for the observed and corrected correlations between the Constructed-response and Forced-choice part of the HCTA. The strength of these correlations urges further research on the extent to which they both measure different aspects of CT. Such a study could compare the strengths of the relations between each part of the HCTA and dispositions associated with critical thinkers, such as tolerance for ambiguity, openness and conscientiousness.

Construct validity on the level of the items of the test could not be assessed or confirmed. The results of the HCTA were not suitable for a PCA. The MIRT model analyses for the CCTT suggested that a unidimensional model fits better than a five-dimensional model. This finding indicates that the five separate scales correlate sufficiently highly to form a single dimension. The sample size of the current study was not sufficient to estimate a confirmatory model with five dimensions with a correlated factor structure. Another explanation for the better fit of the one-dimensional model above the five-dimensional model are the low reliabilities of the subscales of the test. The latter may also be an explanation why the results of the HCTA were not suitable for PCA.

The medium to strong correlations between the November and May scores on the same test are a positive indication of the criterion validity of both tests. The results of the May administration can be predicted based on the November scores.

With the HCTA a growth in CT was assessed, while no difference was found between both administrations of the CCTT, although there was sufficient possibility for growth.

6.4.3 Feasibility and Attractiveness

Concerning feasibility and attractiveness, both instruments have their own strengths. The CCTT surpasses the HCTA with regard to feasibility. It took considerably less time to administer and to score. This difference was mainly due to the use of constructed-response questions in the HCTA. With regard to the test administration, there were more students who thought the HCTA took too long to fill in than students who thought the CCTT took too long. This was linked to the amount of writing students had to do, as students indicated the quantity of writing as a disadvantage of the HCTA. With regard to the scoring, the Vienna testing system was practical, but nevertheless scoring remained more time consuming than the scoring of the forced-choice questions. On the other hand the use of constructed-response questions was one of the main strengths of the HCTA when it comes to content validity.

With regard to the attractiveness of both instruments, the students expressed that they preferred the HTCA above the CCTT, because the situations described in the test were more related to their daily-life experiences and because they could express their own thinking more. Students considered both tests as difficult, but not too difficult. However, 40 percent claimed that the CCTT was at least slightly too difficult, compared to only one quarter for the HCTA. This finding may explain partly why still 50 percent of the students thought the CCTT took too long, although it took considerably less time to fill in than the HCTA.

Summarising, the present study showed that none of the two investigated instruments is sufficient in validity and reliability for freshmen in educational sciences. The fact that only freshmen of one major were assessed is a plausible explanation for the different results in comparison to previous research on both tests. Although this is a limitation of the study, it could also point out that the tests might not be suited to study effects of CT interventions within one programme. The translated instruments hold some promising features, but adaptations in order to increase reliability and construct validity are required.

Moreover, the current study indicated that the constructed-response items of the HCTA are both the most appealing and content relevant characteristic as well as the major challenge of the test. The strength of the correlations between the CCTT and the constructed-response items and the forced-choice items of the HCTA call for additional research on the question about the precise relation between CT skills and dispositions and whether the HCTA is capable of assessing both independently.

Chapter 7: Relationship Between Research Integration and The Development of CT[10]

The empirical grounds for the assumed positive effect of research integration on the development of critical thinking (CT) are limited. Therefore, this study aims at empirically investigating differences in CT development between first year students in various programmes, and whether these differences in CT development can be attributed to differences in research integration practices within the programmes. The study compared three programmes on their research integration practices by means of interviews with teachers (N = 27). In addition, a discipline-general CT measure (Scipio) was administered twice to 124 first year students registered in those three programmes, to measure their CT development. The findings showed that programmes differ in the frequency of a diverse set of research integration approaches and research-related learning outcomes. A paired-samples *t*-test revealed that students improved in their CT. However relationships between differences in research integration and the development of CT could not be retrieved. The results caution for assuming simplistic relationships between research integration and development of CT. More clarity is needed on the effect of different research integration practices on student learning before research integration, as a unified concept, is recommended as a panacea.

[10] This chapter is submitted for publication [Verburgh, Schouteden, Evens, Tiruneh, & Elen (2013). *Development of critical thinking in the first year of higher education: The influence of research integration*. Manuscript submitted for publication.].

7.1 Introduction

The call to strengthen the role of research in undergraduate programmes, both within the curriculum and as an extra-curricular activity, is widespread in higher education (Boyer Commission, 1998; Brew, 2006; Heggen et al., 2010; Trowler & Wareham, 2008; Visser-Wijnveen et al., 2012). The ideal of von Humboldt of higher education as a place in which research and teaching are united, by means of the common pursuit of knowledge by the teacher and the student (Simons & Elen, 2007) echoes in higher education research and policy (Baxter Magolda, 2004; Brew, 2006; Heggen et al., 2010). The Boyer commission (1998) reproached American research universities to inadequately prepare undergraduate students for the challenges of professional life or graduate study and proposed to stimulate student learning through inquiry learning and research opportunities. Similarly, the Higher Education Academy proposed to include research-based study in the undergraduate curriculum (Levy & Petrulis, 2012; Trowler & Wareham, 2008). The OECD (1998) argues that all students in higher education would benefit from learning in a research and scholarly culture and that short inquiry projects to solve practical problems are a good way to introduce students to research and critical inquiry. The European University Association (2007) considers research-based education as a strength of European Universities and states that students need to gain research experience and develop research-related skills.

Central to the recommendations to strengthen the role of research in the undergraduate curriculum is the idea that all students -irrespective of their career prospects- can benefit from research integration (Brown & Mc Cartney, 1998; Heggen et al., 2010; Levy & Petrulis, 2012). Research integration would not only be beneficial for learning specific research skills but also for acquiring competences that transcend research. One of these competences is the development of critical thinking (CT) (Heggen et al., 2010; Lopatto, 2009). Being involved in research is considered as "a scholarly process for learning how to define problems and map a line of investigation. It is also a way to induce CT and develop inquiring minds" (Clark, 1997, p. 251). The development of CT is a central goal in higher education (Halpern, 1998; King & Kitchener, 2004; Tapper, 2004; Tsui, 2001). The ability to think critically is considered to be essential in active and engaged citizenship (Kuhn, 1999; Moore, 2004) and crucial to function in complex contemporary society (Barnett, 1990; Bok, 2006). CT can be defined as the ability to make reasoned judgments about ill-defined problems (King & Kitchener, 2004). It entails "interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based" (Facione, 1990, p. 3).

Studies on CT development in higher education revealed that there is a growth in students' CT during higher education, but only to a limited extent and mainly during the first (two) year(s) (Arum & Roksa, 2011; Astin, 1993; Bers et al., 1996; Giancarlo & Facione, 2001; Hagedorn et al., 1999; Miller, 1992; Saucier, 1995). Arum and Roksa (2011), for example, found an average increase of only .18 standard deviations during the first two years of higher education and a significant proportion of students showed no growth at all. Research on epistemic reasoning, which is closely related to CT but with a stronger focus on the reflection on the nature of knowledge (Brabeck, 1983; Chan et al., 2011), revealed similar results. Students grew in their epistemic reasoning during college, but at a slow pace, with modest gains during the first two years (Wood & Kardash, 2002). Moreover, the highest levels of epistemic reasoning seemed to be attained only by advanced doctoral students or during work experiences after graduation (Baxter Magolda, 2004; King & Kitchener, 2004).

Despite limited CT development, the literature on CT is optimistic about the possibility for students to learn to think critically with appropriate instruction

(Halpern, 1998; Tsui, 2002; Valanides & Angeli, 2005). The literature on CT and on research integration are basically separate tracks, although in both, research integration and research involvement is considered as a valuable way to develop CT (Astin, 1993; Clark, 1997; Lopatto, 2009; Tsui, 2002). However, research on precise benefits of research integration for student learning is in its infancy (Heggen et al., 2010; Seymour et al., 2004). Very often benefits on student learning are stated without adequate empirical support (Seymour et al., 2004). Most research on student learning through research integration is concentrated on "Undergraduate research" (Lopatto, 2009; Seymour et al., 2004). Undergraduate research can be defined as "an inquiry or investigation conducted by an undergraduate student that makes an original intellectual or creative contribution to the discipline" (Council on Undergraduate Research (CUR.org)). Studies on undergraduate research indeed showed that students who were actively involved in a research project developed in CT (Kuh, 2007; Lopatto, 2009; Hunter et al., 2007; Seymour et al., 2004). Similarly studies on epistemic growth found indications that experience with research involvement contributes to epistemological development (Rauckhorst et al., 2001; Ryder, Leach, & Driver, 1999; Samarapungavan, Westby, & Bodner, 2006).

However, because of two main reasons it is unclear if these findings of undergraduate research also apply to research integration within the first year. First, a confounding factor in the effect of undergraduate research on CT is the selection of students (Lopatto, 2009). Undergraduate research is mostly extra-curricular, implying that it is not part of the standard curriculum (Lopatto, 2009, Seymour et al., 2004). Possibilities for undergraduate research are often competitive and intended for students who earned already some credits. So, students often choose and/or are chosen to participate in undergraduate research. This might imply that students in undergraduate research differ from students in regular first year modules.
Second, research integration is not a unified concept (Spronken-Smith et al., 2012). Research integration is understood differently by different actors (Annala &

Mäkinen, 2011). Although there is discussion about the breadth of what counts as research integration (e.g., Zimbardi & Myatt, 2012), it is generally accepted that research integration is broader than undergraduate research. Different categorisations of research integration practices exist, based on the nature of the activities students engage in or on the intended learning outcomes (e.g., Healey, 2005; Neumann, 1992; Verburgh et al., 2013a; Visser-Wijnveen et al., 2012; Zimbardi & Myatt, 2012). Different research integration practices might therefore induce different learning activities with different effects on the development of CT.

Currently the empirical grounds for the assumed positive effect of research integration within the curriculum on the development of CT are limited. Therefore, this study aims to empirically investigate the relationship between research integration and the development of CT in ecological settings. This study particularly investigates whether there are differences in CT development between first year students in various programmes, and whether those differences in CT development can be attributed to differences in research integration practices within the programmes. The research questions are:

1. What are the differences and similarities in research integration practices among selected first year programmes?
2. Do students develop in their CT during the first year of the undergraduate curriculum?
3. Can differences in CT development be related to differences in research integration practices?

7.2 Method

7.2.1 Overall Design

The study combines the results of two separate but related data sets collected in ecological settings: one on CT and one on research integration. Data were collected from students and teachers involved in three first year programmes, of three university colleges. "Teacher" is used to refer to the person who teaches the students. The study on CT adopted a pre-test post-test design in order to identify CT development. For research integration, data were collected by means of interviews with the teachers of the modules in the participating first year programmes.

In the setting for this study, Flanders, Belgium, programmes are structural entities leading to a degree, e.g., the bachelor's programme in geography. Modules are smaller units within a degree programme. Upon registration at an institution, students select a specific programme. Modules are organised in years of 60 credits, comparable to one full year of study. In the first year students have to follow a fixed set of modules.

The three programmes were deliberately chosen from the same disciplinary area: hard sciences (Biglan, 1973). The choice for the same disciplinary area was made in order to ease the interpretation of the results because there are indications that research integration might differ between disciplines (Coate et al., 2001; Colbeck, 1998; Rowland, 1996, Trowler & Wareham, 2008).
In addition, there are indications that students differ in their CT development according to discipline (Astin, 1993; Giancarlo & Facione, 2001; Gadzella & Masten, 1998; King & Kitchener, 2004). The choice for hard sciences was based on the willingness of different programmes to participate in the study.

7.2.2 **CT**

Procedure

Students were asked to complete an online-CT test at the beginning of the academic year (October) as a pre-test and again at the end of the academic year (May) as a post-test. In both moments the test was administered during regular teaching time. The total number of students who participated both in the pre- and post-test was 124 (Table 7-1). The dropout between the pre- and post-test was high, especially for programme 3.

Table 7-1

Number of Participants over Different Administration Moments

	Participation		
Programme	Pre-test only	Post-test only	Pre- and post-test
1	11	4	23
2	17	1	42
3	90	7	59
Total	118	12	124

Instrument

To measure CT, the Scipio was used. The Scipio is a discipline-general CT-test developed in previous research with the intention to have a direct measure of CT with a high content validity (Verburgh, François, et al., 2013; Verburgh, François, & Elen, 2012). A discipline-general test was chosen in order to avoid the influence of discipline-specific knowledge. A direct measure of CT was chosen as it provides more reliable results on student learning (Justice, Rice, & Warry, 2009). The test fits well with how CT is understood in the higher education context of the study (Verburgh, François, et al., 2013). It is an integration of two tests translated into Dutch: The Cornell Critical Thinking Test (CCTT) (Ennis et al., 2005) and the Halpern Critical Thinking Assessment (HCTA) (Halpern, 2012). The Scipio consists of 25 questions, with constructed response items (n = 13) and forced-choice items (n = 12). Two items are based on the “Meaning and fallacies” and “Identification of assumptions” scales of the CCTT. The other items are a selection from the HCTA. Each item starts with a description of a problem in a daily life situation. Some of these daily life situations are presented twice, first as a constructed response item, and then with a forced-choice item. For the scoring of the constructed response items there was an extended code book. The inter-rater reliability is high; the weighted Cohen's κ ranges between .70 and .98 (Verburgh et al., 2012). The maximum score on the Scipio is 95. The internal consistency of Scipio on the pre-test is relatively low, Cronbach's α = .61. The internal consistency on the post-test is higher, Cronbach's α =.67 (Cohen, 1988).

Analyses

In order to test whether drop out was selective an independent samples *t*-test was conducted. It was found that the scores on the pre-test for students who participated only then (M = 50.90, SD = 7.49) did not differ from those of students participating twice (M = 50.93, SD = 7.90), $t(240) = -.029$, $p > .05$ (r =.002). In addition, there are indications that there is no difference in the scores on the post-test between students who participated only on the post-test (M = 51.42, SD = 11.48) and those participated twice (M = 52.06; SD = 8.71), $t(134) = -.236$, $p > .05$ (r = .02). However, this result should be interpreted cautiously because of the low number of students who participated only in the post-test.

Development of CT was measured with a paired-samples *t*-test. To study the differences between programmes, an ANCOVA was initially planned with programme as the independent variable, the score on the pre-test as the covariate and the score on the post-test as the independent variable. However, there was no independence between the covariate and the independent variable, so ANCOVA was not suited (Miller & Chapman, 2001). Therefore, an ANOVA with the growth scores as dependent variable was conducted (Maxwell & Delaney, 2004).

7.2.3 **Research Integration in the Programmes**

Procedure

During 27 individual interviews teachers described their first year module(s) ($N = 36$). During the interview the teachers explained what happened in their classes, what they did during the contact moments with their students, what they expected from them during and in between the contact moments, what they wanted their students to be able to do after finishing the module. Because teachers interpret research integration in a specific way (Neumann, 1992; Zamorski, 2002), there was no specific focus on research integration in the opening question of the interview. This more open approach was adopted in order to avoid that teachers would only talk about what they consider as research integration. The interviewer asked clarifying questions to gain enough information. Interviews lasted between thirty minutes to one hour. In case teachers were responsible for more than one module, all modules were discussed during the same interview.

Research-integration approaches and research-related learning outcomes

Based on an analysis of existing categorisations of research integration, Verburgh et al., (2013a) discerned between two identifying characteristics of research integration practices. Research integration can refer to the way students are brought in contact with research and research-like activities. These different ways are labelled "research integration approaches" (RI-approaches). Research integration can also refer to what students are expected to learn in relation to research. These are labelled as research-related learning outcomes (RR-learning outcomes). Verburgh et al. (2013a) identified eight RI-approaches (Table 7-2) and seven RR-learning outcomes, two of which have specifications (Table 7-3). To improve the readability the RI-approaches will be underlined, the RR-learning outcomes will be italicised.

Table 7-2

Eight RI-approaches(Verburgh et al., 2013a)

RI-approach	Description
Facts	Students are confronted with descriptions of research results as if they were facts. The research background of the results is not mentioned. This approach can be experienced in various settings, e.g., attending a lecture, listening to an explanation of a theory by means of examples, watching a documentary film, reading a text or explaining a given text to peers.
Scientific facts	Students are also confronted with research results. Here the research background is touched but no further explanation of the research is given. This approach can be experienced in similar settings as *Facts*, but with touching the research background. For example during a lecture or a documentary film the speaker says that research has revealed these findings or that the name of a researcher or the scientist is dropped, or with in-text references grounding the statements made.
Research-based facts	Students are confronted with research results and the research behind those results. Explaining the results of a study and why they are relevant for the theory under study is an example of this approach, as well as a teacher showing a documentary film exemplifying the experiment underlying the theory being studied.
Research methods	Students are confronted with research methods. E.g., they receive an explanation about an analysis technique and how it should be used or a teacher demonstrating how to handle a machine to do measurements.
Segments-outcomes relevant	Students are actively involved in doing one or more research steps. The research activity is not oriented towards advancing the discipline but it is relevant for the learning of the student. Therefore it could also be called “research-like activities”. This approach can be experienced in diverse settings, e.g., during lab sessions students have to repeat a classic experiment, students receive a question and they have to think about a method to investigate the question or students have to look up information about a given topic.
Segments-outcomes functional	Students are involved in doing one or more research steps. The results are here not only relevant for the students but also intended to add to the advancement of the discipline. E.g., students have to fine-tune a distillation machine, so that the teacher can use it in her experiments, students have to administer questionnaires of a research study of the teacher, or students have to perform a statistical analysis on real research data.
Full study-outcomes	Students are involved in a complete research cycle and the results of the study are relevant for the student. E.g., students do a small literature review on a chosen topic

relevant	but the results of the review are not important or are already known by the teacher.
Full study-outcomes functional	Students are also involved in a complete research cycle, but in contrast to the previous approach, the results are intended to advance the discipline. Here students become members of the research community. E.g., students have to do a literature review that helps the teacher to develop a hypothesis for a future experiment.

Table 7-3

Seven RR-learning Outcomes, with Specifications of Practical Research Skills and Critical Thinking (Verburgh et al., 2013a)

Basic RR-learning outcome Specification	Description *Example*
Results	Students have acquired knowledge from results of research. *Students acquire knowledge of scientific research on glacier systems.*
Underpinnings	Students have gained insight into methodological and theoretical underpinnings of research. *Students understand the importance of the used research method for the validity of the results.*
Practical research skills	Students have developed particular practical research skills.
Formulating a Research Question	*Students can formulate a question that can be investigated*
Finding literature	*Students can find scientific literature in a database given a question and valid keywords*
Developing a design	*Students can conceptualise a study.*
Collecting data	*Students can observe the behaviour of a psychiatric patient*
Analysing data	*Students are able to do a valid statistic analysis of the given data.*
Formulating a donclusion	*Students can formulate an conclusion based on the question and the data given.*
Report	*Students can write an account of what they have done during the lab session.*
Research attitude	*Students do not accept statements as given but looks for supporting evidence.*
Systematic way of reasoning	*Students can approach a problem in a systematic way in order to come to an answer.*
Competence to be a researcher	Students have the competence to be a researcher. *Students can, starting from a problem situation, make and present an*

	analysis about a particular economic-geographic entity, through the integration of written documents, numerical data, field work and statistical analyses
Critical thinking	Students have a critical attitude towards information, knowledge and knowledge construction.
Critical thinking towards oneself	*Students can critically analyse the own behaviour during an internship according to the theoretical framework used in the module*
Critical thinking towards information	*Students learn to critically interpret historical data and apply this critical sense in the interpretation of the present.*
Conscious about perspective others	*Students are conscious that there are multiple perspectives about one problem.*
Ability to handle uncertainty	*Student can formulate a conclusion based on mixed and incomplete information.*
Curiosity	Students have a curiosity towards evolutions in the discipline
	Students are intrigued by the subject of the module and are stimulated to follow up the developments in the field.
Practice	Student have the ability to use research results in new situations or to see the practical implications of research results
	In order to solve a problem situation, students are able to look for relevant research results and apply them.

Analyses

Each interview was audio taped and transcribed verbatim. It was analysed on (a) RI-approaches and (b) RR-learning outcomes (see Table 7-2 and Table 7-3). This implies that for each module it was analysed what RI-approaches the teacher used and what RR-learning outcomes were aimed at. For a detailed description of the data analysis, see Verburgh et al., (2013a, 2013b).

Then the results of the individual modules were integrated at programme level, in order to see what specific practices were typical for each programme. For each RI-approach the proportion of modules of the programme using that specific approach was calculated. Similarly for each RR-learning outcome the proportion of modules of the programme aiming at that specific learning outcome was calculated. These proportions were compared in the last step of the analysis in order to reveal similarities and differences among the programmes.

7.3 Results

7.3.1 RI- practices

The analysis of research integration practices revealed similarities and differences between programmes concerning RI-approaches (Table 7-4) and RR-learning outcomes (Table 7-5).

The RI-approaches Facts, Scientific facts and Research-based facts were about as frequent in all programmes (respectively about 85%, 50% and 35% of the modules). Full study- outcomes relevant was used in none of the programmes. Acquiring knowledge from *Results* of research and the development of *Practical research skills*, were recurrent RR-learning outcomes in all programmes. *Practice* was in all programmes rarely or not aimed at.

Table 7-4

The Proportion of Modules Applying a Specific RI-approach by Programme

RI-approach	Programme 1	Programme 2	Programme 3
Facts	.92	.80	.86
Scientific facts	.50	.50	.57
Research-based facts	.33	.40	.36
Research Methods	.42	.60	.50
Segments- outcomes relevant	.42	.60	.71
Segments- outcomes functional	.00	.00	.07
Full study- outcomes relevant	.00	.00	.00
Full study- outcomes functional	.00	.00	.07
n	12	10	14

Table 7-5

The Proportion of Modules Aiming at a Specific RR-learning Outcome by Programme

RR-learning outcomes Detailed RR-learning outcomes	Programme 1	Programme 2	Programme 3
Results	1.00	1.00	.93
Underpinnings	.67	.00	.21
Practical research skills	.92	1.00	.86
Formulating a research question	.00	.00	.08
Finding literature	.00	.20	.08
Developing a design	.09	.10	.25
Collecting data	.27	.30	.17
Analysing data	.27	.00	.08
Formulating a conclusion	.27	.20	.25
Report	.18	.30	.58
Research attitude	.27	.30	.17
Systematic way of reasoning	.27	.60	.75
Unspecified	.27	.30	.00
Competence to be a researcher	.00	.00	.14
Critical thinking	.58	.40	.79
Critical thinking towards oneself	.29	1.00	.91
Critical thinking towards information	.43	.75	.18
Conscious of perspective others	.14	.00	.00
Able to handle uncertainty	.14	.00	.00
Unspecified	.43	.50	.00
Curiosity	.33	.40	.00
Practise	.08	.00	.07
n	12	10	14

*Note***.** The proportion of the detailed RR-learning outcomes is based on the number of modules aiming at the overall RR-learning outcome. More than one interpretation per occurrence is possible.

Based on the analyses, the programmes can be described as follows.

In less than half of the modules of programme 1 students were actively involved in research-like activities and these activities were always oriented towards student learning (Segments- outcomes relevant, in 40% of the modules).

Typical for programme 1 was the outspoken attention for the theoretical and methodological *Underpinnings* of research and the moderate attention to stimulate students' *Curiosity* for future developments in the discipline (respectively in 70% and 30% of the modules). In programme 1 a broad range of *Practical research skills* were aimed at, and each module concentrated on one or two specific skills, without one skill being aimed at in a majority of modules. *Critical thinking* was an intended RR-learning outcome in a small majority of the modules and it mostly concerned *Critical thinking towards information*.

The frequent confrontation of students with research methods set programme 2 apart (<u>Research methods</u> in 60% of the modules). In addition, students were frequently involved actively in research-like activities, oriented towards student learning (<u>Segments- outcomes relevant</u>, in 60% of the modules).
Programme 2 devoted considerable attention to the development of *Curiosity (*40% of the modules*). Practical research skills* were aimed at in all modules, and mostly this entailed a *Systematic way of reasoning* in combination with one other practical research skill. *Critical thinking* was aimed at in a minority of modules and it entailed both *CT towards oneself* and *towards information*.

Characteristic in programme 3 was high occurrence of active involvement in research-like activities, of which most were oriented towards student learning (<u>Segments- outcomes relevant,</u> 71% of the modules). In one module students were involved in a complete research study, of which the results were new to the discipline (<u>Full study-outcomes functional</u>).
In programme 3 *CT* and *Practical research skills* were recurrent learning outcomes. *Critical thinking* mainly referred to *Critical thinking towards oneself*. The intended *Practical research skills* were *Report* and a *Systematic way of reasoning*. In one module development of the *Competence to be a researcher* was aimed at.

7.3.2 Development in CT

The paired-samples *t*-test revealed that on average, students got higher scores on CT on the post-test than they did on the pre-test, $t(123) = 2.025$, $p < .05$, with a small effect size ($r = .18$).

Table 7-6

Descriptive Statistics of CT Scores on Both Test Moments.

		Pre-test		Post-test	
Programme	n	*M*	*SD*	*M*	*SD*
1	23	47,00	6.22	48.91	10.45
2	42	50.60	7.66	52.36	8.64
3	59	52.69	8.17	53.07	7.86
Total	124	50.93	7.90	52.06	8.71

7.3.3 Differences in Development in CT between Programmes

There was no significant difference between the three programmes concerning growth in CT, $F(2,121) = .84$, $p > .05$.

7.4 Discussion

The study investigated students' growth in CT in three first-year programmes in relation to differences in the research integration practices.

Students clearly developed their CT during the first year in higher education. However, this development was limited and comparable to CT growth found in other studies, like for example Arum and Roksa (2011). The results also showed that although the three programmes differed in their research integration practices, no significant difference was found in CT development among the students involved in the programmes.

There were some remarkable differences between RI-approaches and RR-learning outcomes, based on which a differential growth in CT could been expected.

First, differences between RI-approaches are discussed. The involvement in a RI-approach comparable to undergraduate research did not result in higher gains in CT. In programme 3 there was one module in which students were involved in an authentic research study (Full study- outcomes functional). The design of the module showed similarities with the features of undergraduate research (Lopatto, 2009): Students had to design some aspects of the project, students had to communicate the results orally and write a report, teachers provided structure and guidance. The experience was deviant from prototypical undergraduate research with respect to time investment (more restricted: 180 hours for the entire module, of which the research project was only one small part) and framing in scientific literature (very limited). Moreover, all first year students of a programme were involved. These findings concur with Lopatto (2009) and Seymour et al. (2004) by showing that the impact of research participation is not straightforward.

In addition, a frequent confrontation with the RI-approach Research methods does not necessarily imply that students improve their CT performance more. Lopatto (2009) argues that the involvement with methodological issues is beneficial for CT. However, in our findings, such an impact is not retrieved. Students in programme 2 were more frequently confronted with research methods than students in the other two programmes but this did not result in stronger CT-growth.
Also differences in the frequency of participation in research-like activities (Segments- results relevant) did not result in a differential growth in CT. In this RI-approach students are actively involved in one or more research steps and the activity is focused on student learning. Lopatto (2009) defined research-like courses as courses that contain some or all of the following five characteristics: 1) In the course there is a project with an unknown answer, also to the teacher; 2) students have some input in the research process; 3) students design their own project; 4) students are responsible for an aspect of the project; 5) students have to critique other students' work. He maked a distinction between high and low research-like courses, depending on the emphasis on the mentioned characteristics. A comparison between

self-reported learning gains (on aspects closely related to research) between students involved in undergraduate research, high research-like, and low-research-like courses, revealed that undergraduate research students reported highest gains, followed by students in the high research-like courses. Students in the low research-like activities reported lowest gains. The modules using Segments-results relevant would predominantly be labelled as low research-like courses. If the learning gains of CT are in line with the learning gains of aspects closely related to research, then the effect of low research-like courses on CT might also be low. This could explain the absence of a differential effect between programmes that more or less apply research-like activities.

Second, differences in RR-learning outcomes seem to have no direct effect on the development of CT either. A difference in the importance of the RR-learning outcome *Underpinnings* did not coincide with a difference in growth in CT. The literature argues that gaining insight in the underlying concepts of research and research results (i.e. what *Underpinnings* is about) could be conducive to epistemic development (King & Kitchener, 2004). Gaining insight in the theoretical and methodological *Underpinnings* of research involves higher order thinking, which is beneficial for development in CT (Halpern, 2001). However, this was not reflected in the results of this study as *Underpinnings* was an important learning outcome in programme 1 in contrast with the other two programmes without a difference in CT development. Similarly, despite differences in the frequency of *Critical thinking* as a RR-learning outcome and its subcategories, there was no difference in growth either. Abrami et al. (2008) concluded in their review on CT teaching approaches that the explicit integration of CT goals in the design of the course is important. Therefore we would expect more growth when CT is an intended learning outcome. Moreover, differences in the kind of *Practical research skills* aimed at, did not seem to have an influence on CT development.

This lack of relationships might be due to a gap between what teachers aim at and the effectiveness of their actual teaching approach (Browne & Freeman, 2000; Murray & MacDonald, 1997). Browne and Freeman (2000) claimed that CT is more common as an educational objective than as an actual encouragement in the classroom. A study of Paul, Elder and Bartell (1997) in teacher training found that the majority of teachers claimed CT to be important but only a small minority could give a clear definition of CT and almost 80% had no idea how to stimulate CT in the classroom. So, it is possible that teachers in the current study had the intention to stimulate CT but their teaching approach was not the most optimal way to do so, which resulted in limited CT development.

However, interpreting of the absence of differences between the programmes is hampered by at least three factors. First, the programmes differed from one another in various respects. Programmes used different RI-approaches and aimed at diverse RR-learning outcomes at the same time. They differed from one another on more than one aspect. Therefore the effect of a one aspect of research integration (prominent in one programme) was possibly counterbalanced by the effect of another (prominent in another programme). Moreover, the study was spread over one academic year, during which students were confronted with many experiences that could influence their CT and were or were not related to research integration. This limitation is inherent to the ecological setting of this study and a feature of reality in higher education.

In addition, the instrument to measure CT was not free of noise, indicated by the rather low Cronbach's α. However, this limited reliability might also relate to the multi-dimensionality of CT (Cortina, 1993).

Third, it is possible that students mainly acquired discipline-specific CT skills which were not tested in a discipline-general test such as the Scipio. There is general agreement among CT researchers that CT consists of discipline-specific and discipline-general CT skills. Discipline-general skills are applicable in various contexts and disciplines (Angeli & Valanides, 2009). It is possible that students mainly developed discipline-specific CT-skills which are not assessed by the Scipio. Research indicates discrepancies between development in discipline-specific and discipline general CT (Stark, 2012). Students show more development on discipline-specific CT and transfer to discipline-general appears to be difficult (Williams, Oliver, & Stockdale, 2004).

Despite the limitations of the study, the results exemplify the warning for falsely assuming simplistic relationships between research and teaching/learning (Hattie & Marsh, 1996; Terenzini & Pascarella, 1994). The study questions the assumption that research integration in the undergraduate curriculum leads to CT development and hence more research should be integrated in the curriculum. The mere fact that research is integrated in the curriculum is not enough to induce a large development of CT. The study also shows that in the research integration practices there is no systematic use of instructional approaches that are stimulating to achieve specific learning outcomes as CT. The modules are not designed to foster CT. Although the results of research on CT instruction are inconclusive (Niu et al., 2013; Pascarella & Terenzini, 2005; Valanides & Angeli, 2005), there are different suggestions about instructional approaches that could effectively develop students' CT. For example, integrating explicit CT instruction within subject-matter content (the *mixed* approach) has been consistently effective in developing CT (e.g., Abrami et al., 2008). Moreover, teaching strategies that includes teacher modeling (Halpern, 1998; Hofer, 2004; Kronholm, 1996); using ill-structured, messy, complex problems and cues in problems that can serve as triggers for applying CT (Halpern, 1998; King & Kitchener, 2004); critiquing a paper, essay exams, collaborative work, class presentation and discussions, writing and rewriting with much attention to analysis, synthesis, and refinement of ideas (Astin, 1993; Tapper, 2004; Tsui, 1999; 2002) have been found effective in developing CT. These teaching strategies can be – and most probably are to a different extent- applied in research integration practices. But the current spontaneous application is most probably not on continuous basis or unsystematic. For example, Samarapungavan et al. (2006) found that the more students had the opportunity to ask questions and discuss with researchers during their research experience, the higher was their epistemological growth. During these discussions researchers probably modelled their thinking to the students. Thiry and Laursen (2011) point at the crucial role of the supervisor for the quality of the learning gains in undergraduate research. They claim that supervising is "an educational activity with specific pedagogical practices that are successful and others that are not" (p. 782).

The lack of systematic design possibly results in missed chances for CT development because a carefully designed learning environment can result in high gains in CT (Niu et al., 2013; Pascarella & Terenzini, 2005). It is likely that the potential for effective research integration to induce CT is not yet fully exploited. A valuable direction for future research on research integration would therefore be to investigate the effectiveness of systematically designed research integration practices on the development of CT.

While most research on research integration is performed in universities (Heggen et al., 2010; Neumann, 1992; Visser-Wijnveen et al., 2012; Zimbardi & Myatt, 2012), this study was conducted in university colleges, which are teaching-intensive institutions, The situation and the rationale for research integration might differ between universities and university colleges (Brown & McCartney, 1998). In general university colleges are higher education settings where doing research is relatively new and universities are higher education settings with a long research tradition, while acknowledging this is an oversimplification of the complexities involved (Bleikie, 2003). In university colleges, teachers are increasingly expected to do research and integrate research into their teaching (Kyvik, 2009; Kyvik & Lepori, 2010; Witte et al., 2008). The prevalence of RI-approaches or of RR-learning outcomes therefore possibly differs between university colleges and universities. A comparison between both types of institutions would be valuable to know more on the understudied impact of the type of institution on research integration (Jenkins, 2004).

However, and in addition to the fact that the difference between the labels universities and university colleges is relative (Witte et al., 2008), there is little ground to assume that teaching at universities is more systematically designed in comparison to the teaching in the university colleges. Therefore, the findings of the suggested design research are also relevant for universities and would help to fulfil larger development in CT in higher education.

Chapter 8: Discussion

8.1 Introduction

This doctoral dissertation focussed on two main questions: (1) What is the prevalence of research integration and (2) What is the effect of research integration on the development of critical thinking. To answer these questions, different studies were conducted as described in the previous chapters.

In this chapter findings are summarised and discussed. The chapter is organised in four main parts. The first -and most complex- part relates to the prevalence of research integration. After a short summary of the main findings on research integration, the impact of the selected RR-learning outcomes and RI-approaches is discussed. Next, the methods to assess research integration are considered. The first part ends with a comparison with the literature and ideas for further research, concerning the prevalence of RR-learning outcomes, concerning the relation between RR-learning outcomes and RI-approaches and concerning defining research integration.

The second part deals with the development of CT. The results are summarised and subsequently critically discussed.

The third part concerns the relationship between research integration and the development of CT. It begins with the main findings of this doctoral dissertation on this issue. Next, possible explanations are given, followed by implications and ideas for further research.

The chapter ends with an overall conclusion.

8.2 Research Integration

8.2.1 **Main Findings**

The first research question pertains to the prevalence of research integration. This doctoral dissertation revealed first of all that research integration is a reality at the level of the selected RR-learning outcomes. It seems that in higher education learning outcomes in relation to research are intended. Moreover, it appears that some of these RR-learning outcomes are more frequently intended than others. This was found when investigating the module descriptions (Chapter 3 on programme patterns) and endorsed when talking to teachers about their first year modules (Chapter 5 on RR-learning outcomes by module type). Although the prevalence of the distinct learning outcomes is different in both studies, identical RR-learning outcomes are most frequently aimed at: *Results*, *Practical research skills* and *Critical thinking*.

Second, teachers confirmed the use of different identified RI-approaches (Chapter 4 on RI-approaches). Some approaches seem to be more frequently used than others. Facts is the most frequent RI-approach, used in the majority of the modules. Furthermore, active involvement in research-like activities is almost exclusively concentrated on activities relevant for student learning (Segments-outcomes relevant). Research for the sake of truth (Segments or Full study- outcomes functional) is hardly used.

Third, programmes differ in the way they integrate research. Three distinct patterns in the prevalence of RR-learning outcomes between programmes were identified (Chapter 3 on programme patterns): A low-attention pattern with an overall low attention for the RR-learning outcomes; a results pattern with a higher attention for the RR-learning outcomes, especially pertaining to acquiring knowledge of research results (*Results*); and a critical-thinking pattern with the main focus on *Critical thinking*, *practical research skills* and the *competence to be a researcher*. The critical-thinking pattern can be considered as the pattern most focused on undertaking research or becoming a researcher. In addition, fine-grained differences between programmes at the level of RR-learning outcomes and RI-approaches were retrieved (Chapter 7 on CT-development by research integration).

Fourth, this doctoral dissertation showed that modules differ in the way they combine one or more RI-approaches and privileged relationships seem to exist between these combination and RR-learning outcomes (Chapter 4 on RI-approaches and Chapter 5 on learning outcomes by module type). In the majority of the modules different RI-approaches are combined. Six different module types were found. A module type is a set of one or more approaches within one module. The module type **Insightful practice (**a combination of five RI-approaches) is most common, followed by **Facts only** (with only <u>Facts</u> as RI-approach) and **Facts plus** (a combination of <u>Facts</u> and <u>Scientific facts)</u>. The privileged relationships between the module types and the RR-learning outcomes were concentrated on *Underpinnings, Critical thinking* and *Curiosity*. In addition, a more fine-grained analysis disclosed that although all modules aim at the development of *Practical research skills*, there are substantial differences in the intended specific research skills. Similarly, the meaning of *Critical thinking* differed along the module types. For the other RR-learning outcomes such privileged links were not retrieved: In all modules *Results* are aimed at, and in none the *Competence to be a researcher* or *Practice*.

8.2.2 Selection of RR-learning Outcomes and RI-approaches

When studying research integration, an important question is what exactly counts as research integration. Opinions differ considerably among teachers, students and policy makers (Durning & Jenkins 2005; Neumann, 1992, 1993, 1994; Visser-Wijnveen, 2009; Zamorski, 2002). But also researchers on the topic hold different opinions. For example, Zimbardi and Myatt (2012) confine their study to teaching approaches that imply active involvement of students in research-(like) activities, while During and Jenkins (2005) also include "Talking about research conducted by the academic". Moreover, sometimes the criteria to label a teaching practice as research-based or research integration are not made entirely explicit, for example in projects on "good practices" of research-based teaching.

In this doctoral dissertation two facets of research integration (RR-learning outcomes and RI-approaches) with a broad interpretation of what counts as a RR-learning outcome and a RI-approach were deliberately selected.

The selection of the RR-learning outcomes and RI-approaches provokes questions about what is meant and aimed at with research integration.

For RR-learning outcomes, for instance, one might argue that the selected ones are too inclusive. Some RR-learning outcomes are general in nature and the specific link with research or research integration is rather vague. When the virtues of research integration for student learning are lauded for instance, it is questionable if solely acquiring knowledge from results of research (*Results*) is considered sufficient. The importance of *Results* is beyond question, but its distinctive nature is questionable, since domain specific knowledge is important for many types of learning (Glaser, 1984) and domain specific knowledge most commonly refers to results of research. In addition, a *Systematic way of reasonin*g (as a *Practical research skill*) might also be not very illuminating with respect to research integration. In some cases the relationship between this RR-learning outcome and research is obvious, for example when teachers want their students to learn to analyse a problem situation in a systemic way in order to come to a precise formulation of the problem. However this relationship is often far less clear, for instance when teachers want their students to learn to think thoroughly. In our analyses, such learning outcomes were given the benefit of the doubt. The same counts for including *Critical thinking towards oneself*. The relevance of the learning outcome is pertinent and also van der Rijst (2009) relates a "Self-critical attitude" with a "Scientific research disposition". It is doubtful however whether this thinking always surpasses the particular and whether it is always systematic and based on theoretical or conceptual considerations, features considered to be preconditions to label an activity as research (Bills, 2004; Neumann, 1993).

At the same time, one could also argue that the selected RR-learning outcomes are not exhaustive. Discussions with representatives of different institutions of higher education about the results of the study on programme patterns (Chapter 3) and about reasons to integrate research, suggested at least one missing RR-learning outcome (Elen et al., 2011): the ability to apply research results to solve practical problems or to identify the relevance of particular research results for the profession. This ability was considered as an important aspect of research integration but absent in the original six RR-learning outcomes (Elen et al., 2009) used in the study on programme patterns. Therefore, *Practice* was added as an extra RR-learning outcome (Elen et al., 2011) and studied in the study on RI-approaches (Chapter 5). However, *Practice* seemed to be rare in the investigated first year modules.

Similarly for the RI-approaches one could on the one hand argue that they are insufficiently selective. The eight studied RI-approaches cover a broad spectrum of research integration. Teaching practices that involve research-like activities with results relevant for students, but that are not oriented towards advancement for the disciplines (Segments/Full study- outcomes relevant) are included. This position is also taken by Brew (2010) and Levy and Petrulis (2012), but not by Beckman and Hensel (2009). Moreover, also approaches in which students are confronted with research without doing research(-like) activities are included. This is similar to "Students as audience" of Healey and Jenkins (2009), but more inclusive than the models of Zimbardi and Myatt (2012). When the claim is made that research integration is a distinctive characteristic of higher education, that sets it apart from other types of education like secondary education, it is disputable whether one has a

RI-approach as <u>Facts</u> in mind. The substance of many other types of education, is also based on research and teachers will confront their students with that content. Hence, it is most probably that also in other types of education, <u>Facts</u> will be a common RI-approach.

On the other hand a further elaboration of the RI-approach <u>Segments-outcomes relevant</u> could be considered. In this doctoral dissertation this approach entails quite diverse practices, such as cookbook lab-sessions as well as assignments to develop a research design to test a given hypothesis.

8.2.3 **Methods to Measure Research Integration**

This doctoral dissertation attempted to grasp actual practices concerning research integration. Therefore concrete learning environments are taken as point of departure. Although departing from cases studies or concrete practices (e.g., Elsen, et al., 2009; van der Rijst, 2009; Visser-Wijnveen et al., 2012; Zimbardi & Myatt, 2012) is common, a lot of research on the topic starts from teaching in general. Those studies look for example for possible instances of research integration (During & Jenkins, 2005; Neumann, 1992; Robertson & Bond 2001, Zamorski, 2002). The perspective taken in this doctoral dissertation implies that only part of teachers' full register of possibilities is studied. The study is confined to the practices of the modules under study.

In order to grasp actual teaching practices concerning research integration, the applied methods – analysis of the module descriptions and interviews – both depart from all-round descriptions. The module descriptions serve multiple purposes (Parkes & Harris, 2002). The interviews departed from the general question to describe what happened during contact moments with students. In the opening question there was no explicit focus on research integration, although the interviewer made sure that research integration – as understood by the interviewer- came to the surface. Based on those descriptions it was the researcher who decided for each of the RR-learning outcomes and RI-approaches if it was present or not, for teachers interpret research integration in different ways (Neumann, 1992; Zamorski, 2002). This method has as advantage that the impact of the diversity in interpretation gets diminished.

However, there is possibly a difference between teachers' words and teachers' behaviour. Module descriptions pertain to the formal curriculum, which does not necessarily mirror the enacted curriculum (Marsh & Willis, 1999). Teachers may add as well as dismiss formally established learning outcomes. Moreover, module descriptions tend to be compressed and fuzzy, allowing multiple interpretations (Hrycaj, 2006). The interviews are intended to shed light on the enacted curriculum. Interviews build on the assumption that teachers are able and willing to accurately describe their aims and approaches. The relation between teacher thinking and teacher practice is however complex (McAlpine, Weston, Timmermans, Berthiaume, & Fairbank-Roch, 2006). What teachers tell they do may therefore differ from the enacted curriculum. There are indications that it is difficult for teachers to talk in a precise way about their teaching (Mälkki & Lindblom-Ylänne, 2012). Observation studies could therefore be more appropriate to study the enacted curriculum. This can be exemplified with a study of Reisman (2013). In her study, different teachers implemented the same carefully designed curriculum to stimulate students argumentation skills. Through class observations she revealed important differences between teachers on the profundity of the argumentation during classroom discussions. When Reisman would have talked to teachers about their practices, she might not have found these differences because teachers might all have described their acting as "Stimulating a discussion".

Similarly, what teachers write or tell about their teaching practices could differ from students' experiences of the curriculum. In the model of Terenzini and Reason (2013) (Figure 1-1), curriculum experiences are the received curriculum (Kelly, 2009). Students' experiences are related but not equal to the enacted curriculum. A study of Visser-Wijnveen (2009), for instance, showed that the intended learning outcomes stated by teachers are more modest than what students report to have learned, especially in relation to awareness of research and research disposition. Students may perceive the development of certain competencies, such as the *Competence to be a researcher,* as a learning outcome even when the teacher has not stated it as an explicit learning outcome. Van der Rijst (2009) also found differences between teachers' intentions and student perceptions of those intentions.

When interpreting the results, it is important to keep in mind that only presence of the RI-approaches or the RR-learning outcome is assessed. A module using a RI-approach in only a minor part of it and a module using a RI-approach intensively, are equally considered as modules using that RI-approach. Similarly, the relative importance of learning outcomes within a module is not taken into account. Moreover, neither the complexity nor the level of difficulty are taken into account. For example acquiring basic knowledge on the research on topic X would receive the same score for *Results* as acquiring advanced knowledge on the research on topic X in relation to topics Y and Z.

Although both methods to assess the RR-learning outcomes share the same points of departure, learning outcomes written in module descriptions and those mentioned in interviews possibly differ because interviews were more focussed on the topic of research integration than the module descriptions.

Our data permit to investigate this potential discrepancy since there is a module description available for all modules discussed during the interviews. To make this comparison, all module descriptions were first analysed according to the key used in the study on programme patterns (Chapter 3), complemented with the additional RR-learning outcome *Practice.* The threefold scale (No indications, Vague indications and Clear indications of the pursuit of the learning outcome) was used. Then scores were recoded to the dichotomous scale used for the interviews (RR-learning outcome aimed at or not aimed at), with Vague indications becoming "RR-learning outcome aimed at". The analysis revealed that both methods have about the same proportion of modules aiming at a particular learning outcome, except for *Underpinnings* for which the module descriptions had a higher proportion (Table 8-1). The proportion equal codes between the two methods indicates whether both methods identify the same modules as (not) pursuing a learning outcome. Chi-squared analyses could not be performed because too often the expected value in the cells was lower than five. The analyses showed high proportions equal codes for *Results*, *Practical research skills*, *Competence to be a researcher* and *Practice*. For the other three RR-learning outcomes (*Underpinning*, *Critical thinking* and *Curiosity*) the proportion was low, indicating differences between the two methods in which modules were identified as (not) pursuing the learning outcome.

Table 8-1
Comparison between the Results of the Module Descriptions and the Interviews (N = 46[11])

RR-learning outcome	Proportion modules aiming at a learning outcome*		Proportion equal codes
	Module description	Interview	
Results	.91	.98	.89
Underpinnings	.54	.35	.52
Practical research skills	.89	.94	.87
Competence to be a researcher	.09	.04	.87
Critical thinking	.67	.67	.52
Curiosity	.28	.28	.57
Practice	0	.04	.96

Note. The scores of the module descriptions were transformed to a two-point scale (RR-learning outcomes aimed at or not aimed at) similar to the scale used for the analysis of the interviews.

Proportions equal scores could be low for different reasons. First, the learning outcomes *Underpinning*, *Critical thinking* and *Curiosity* is possibly more related to the

Chapter 8

[11] The total number of modules here (N= 46) is high than in the RR-learning outcomes in module types study (Chapter 5) (N=31) because there only the modules belonging a module type were included.

intangible nexus, while *Results* is more related to the tangible nexus (Neumann, 1991). The intangible nexus pertains to the subtleties of the research process and to the development in students of an approach and attitude towards knowledge, whereas the tangible nexus relates more to the acquisition of scientific findings. The intangible nexus is more difficult to grasp and to clearly define. The breadth of interpretation of *Underpinning*, *Critical thinking* and *Curiosity,* is larger and therefore harder to clearly distinguish. Whereas in interviews interpretation differences can be further explored, this is not possible for written phrases only. Visser-Wijnveen et al (2010) and van der Rijst (2009) also reported difficulties to comprehend those intangible aspects of research integration.

Closely related to the difficulty to grasp *Underpinning*, *Critical thinking* and *Curiosity,* is the difficulty to establish a high interrater reliability for *Underpinnings* (Weighted Cohens' κ = .61) in the scoring of the module descriptions. This difficulty was not experienced for *Critical thinking* and *Curiosity* given the high interrater reliabilities (respectively weighted Cohens' κ = .88 and .83).

Another possible explanation for low equal proportions could be that teachers mention more learning outcomes during the interviews than actually written down because they aim at those learning outcomes without assessing them. The difference between the learning outcomes mentioned and assessed could be less prominent in the module descriptions because in most programmes there is the policy that learning outcomes mentioned in the module descriptions should be assessed. However, our data do not support this explanation, as the proportion modules aiming at specific learning outcomes is not higher for interviews in comparison to module descriptions.

A fourth possible reason is the practice of using standard phrases in the module descriptions. In some institutions, teachers receive a list of standard phrases of learning outcomes from which to select those they aim at. If such a phrase implies a combination of learning outcomes (such as a combination of *Results* and *Underpinnings* or *Practical research skills* and *Critical thinking*) a teacher could select the phrase for only one of them. If that is the case the module description method would find indications for both learning outcomes, whereas the interview method would find indications for only one.

Despite the rather good match between the two methods on the proportion of modules aiming at each of the seven learning outcomes, the low proportion of equal scores for *Underpinning*, *Critical thinking* and *Curiosity* calls for further research. The two methods found indications for each of these RR-learning outcomes in different modules which raises questions about their validity.

8.2.4 **Comparison with the Literature and Ideas for Further Research**

Prevalence of RR-learning Outcomes

The study on programme patterns (Chapter 3) was one of the first to analyse the prevalence of RR-learning outcomes on a larger number of programmes in comparison to other more case-based studies (van der Rijst & Jacobi, 2010; Sin, 2010). The comparison between the prevalence of RR-learning outcomes in the study on programme patterns and the prevalence in the first year modules of Data collection 3, reveals a higher prevalence of *Underpinnings* and *Curiosity* in the first year modules and a lower prevalence of the *Competence to be a researcher*.

These differences could be related to the fact that in the latter study only first year modules were included whereas the study on programme patterns concerns full programmes. There are indications for an increase in the level of research integration into teaching as students advance in their study (Neumann, 1992). Teachers stated to encounter more difficulties in integrating research into their first year classes in comparison to the later years, as first year students lack basic knowledge, maturity or specific critical capacities (Elen & Verburgh, 2008; Neumann, 1992; Taylor, 2007). This could result in differences between the prevalence of the different RR-learning outcomes in the first year and the overall programme.

Overall, the study on programme patterns and the differences with the first year modules call for further research on the stability of the retrieved patterns in a larger sample of programmes and between different years of study within programmes. The refinements *Practical research skills* and *Critical thinking* could be taken into account as well as the relation with *Practice*.

A relevant factor to investigate within this larger sample is the effect of universities versus university colleges or the effect of academic programmes versus professional programmes. Most research on research integration was done in university settings (Heggen et al., 2010) and the effect of institute on research integration is understudied (Jenkins, 2004). In the study on programme patterns (Chapter 3) university and university college academic programmes were studied together with university college professional programmes. The study revealed that despite different labels, the traditional distinction between professional and academic programmes was not clear-cut. This seems in line with the claim of an academic drift in university colleges (Lepori & Kyvik, 2010). It also concurs with Verhoeven and Beuselinck (1996) that the traditional distinction between professional and academic programmes is relative because also some academic programmes have a strong preparation for a profession, e.g., Master's programme of medicine to become a physian.

In identifying programme patterns no distinctions between different specifications of *Practical research skills* and *Critical thinking* were made.

This possibly masked differences between universities and university colleges or between academic and professional programmes. Given variability in the frequencies of the specifications of *Practical research skills* and *Critical thinking*, it could be illuminating to study whether different programme patterns denote these learning outcomes differently, in relation to the type of institution. For instance, the specific *Practical research skills* aimed at in the Results pattern possibly differ from those aimed at in the Critical thinking pattern. Similarly, differences in the interpretation of *Practical research skills* and *Critical thinking* could exist between universities and university colleges, or between professional and academic bachelor's programmes. Such differences would imply that, although teachers use the same words, they do not necessarily share the same interpretation or aim at the same learning outcome. Insight in the underlying interpretations of RR-learning outcomes is important as it appears to have an influence on teaching approaches (Van Hertbruggen, 2013).

Upon request of higher education representatives *Practice* was added as a RR-learning outcome. Since this RR-learning outcome was not studied in the study on programme patterns (Chapter 3), it is currently unclear how the three identified patterns relate to this additional learning outcome. Based on the absence of *Practice* in most of the investigated first year modules (Chapter 5), it is difficult to make precise predictions. It is possible that *Practice* becomes more prominent in advanced modules or at universities.

Relationship between RR-learning Outcomes and RI-approaches

Module types and their relation to learning outcomes, show resemblances and differences with the findings of Visser-Wijnveen et al. (2010, 2012) and to a lesser extent with those of Zimbardi and Myatt (2012). Most module types found could be linked with those of Visser-Wijnveen and her colleagues (2010; 2012). Module types like **Facts research-like illustrated** and **Research process explained and applied** were not retrieved in Visser-Wijnveen's study, but they resemble the "Methods course" of Zimbardi and Myatt (2012), although the latter is probably more in-depth. An important difference is that in this doctoral dissertation only few modules in which students participate in authentic research were found (Segments or Full study-functional). In addition to different inclusion criteria, this doctoral dissertation included only first year modules. This might be an important difference with other studies as van der Rijst and Jacobi (2010) found only a limited amount of authentic research activities in bachelor's programmes.

The limited participation in authentic research (Segments or Full study-functional) could also relate to the type of institution. Whereas the literature on research integration is largely based on research in universities, this doctoral dissertation focused (although not exclusively) on university colleges. Teachers in university colleges may see less opportunities to involve students in authentic research. Until recently research in these institutions was the personal initiative of the individual teacher (Verhoeven, 2010). Griffioen (2013) points at doubts of teachers in university colleges about their own research capacities which influences their willingness or confidence to engage in research activities. Moreover, students in university colleges expect more than their fellow students at the university to be trained for a job (De Wit & Verhoeven, 2003). Therefore teachers possibly think that their students are not interested in being involved in research.

Moreover, the results of the study on RR-learning outcomes in module types (Chapter 5) also indicate the value of looking at module types and not simply at individual RI-approaches. Although module types share the same RI-approach, they do not necessarily share the same RR-learning outcomes. Combinations of RI-approaches appear to be a more functional unit of analysis to study research integration than individual RI-approaches, as also shown in the studies of Visser-Wijnveen (2010; 2012) and Zimbardi and Myatt (2012). The description of educational practices at module level will probably be more informative than a more analytical description at the level of single approaches.

In addition, the results indicate that a multitude of RI-approaches within one module is not necessarily desirable. The use of more RI-approaches does not imply the pursuit of more RR-learning outcomes (The study on RR-learning outcomes in module type, Chapter 6). If the pursuit of RR-learning outcomes is desired, this doctoral dissertation pleads for well-thought-out learning environments with careful considerations about the fit between RI-approaches and RR-learning outcomes. A lean learning environment could help to simultaneously aim at more learning outcomes.

Future research on RR-approaches in combination with RR-learning outcomes in more diverse settings could help to reveal the stability of the identified module types and their relation with RR-learning outcomes. The comparison with the module types of Visser-Wijnveen (2010; 2012) and Zimbardi and Myatt (2012) indicates some possible discrepancies. Moreover the retrieved relationships between RR-learning outcomes and RI-approaches are far from self-evident and therefore require further study.

What is Research Integration?

Irrespective from the opinion about the selection of RR-learning outcomes and RI-approaches being too inclusive, the study calls for defining research integration, for making explicit what is really aspired for and how the actual learning environment looks like. Discourse on the teaching-research and research integration is omnipresent in the literature as well as in daily practice. However confusion and fuzziness about the current approaches and their intended learning outcomes remains due to limited vocabulary and negligent use of existing concepts (Trowler & Wareham, 2008; Spronken-Smith et al., 2012). This call for precise language applies to teachers and policy makers, and to researchers as well.

In Flanders, the setting of this doctoral dissertation, important changes in the landscape of higher education are currently taking place, as the academic programmes of the university colleges will be integrated into the universities and for both programmes the same evaluation criteria in governmental quality assurance initiatives apply (Integration Decree, 2012; Verhoeven, 2010; VLUHR, 2013). During the academisation process (Verhoeven, 2010), external quality assurance paid attention to "the interaction between research and teaching in the curriculum and the corresponding requirements for personnel" [Own translation] (VLIR-VLHORA, 2007,

p. 17). From 2013 onwards attention is devoted to the developed policies of the institutes to "support the programmes based on the mandates concerning research and concerning public and scientific service" [Own translation] (VLUHR, 2013, p. 5). This explanation can be interpreted in diverse ways and does not offer clear guidance nor criteria. It could be valuable for programmes to know better what is expected concerning this issues in the academisation process.

Moreover, research on the topic of research integration would benefit from precise conceptualisations. The results indicate that precision and attention for details add to our understanding of the complexities involved. Distinctive differences in what teachers aim at are more likely to be found at a fine-grained level than at a more general level, as shown in the differences in detailed interpretations of *Practical research skills* and *Critical thinking* between different module types. The detailed distinction between RI-approaches made it possible to identify module types, which appear to be linked to specific combinations of RR-learning outcomes. Precision is needed to improve our understanding of research integration.

The selections of RR-learning outcomes and RI-approaches contribute to the development of a clear vocabulary to talk about the vexed concept of research integration in a precise way.

8.3 CT

8.3.1 Main Findings in Relation to Expectations

The results of this doctoral dissertation indicate that students develop in their CT, but to a limited extent. This is reflected in significant higher scores on the CT-test at the end of the academic year in comparison to the beginning. In the study on CT-instruments (Chapter 6) the average growth on the HCTA was 4.25 (SD = 10.27, $t(154) = 5.14$, $p < .001$, $r = .38$). For the CCTT a significant growth was not established. In the study on research integration and CT development (Chapter 7) the growth was 1.13 (SD = 6.21; t (123) = 2.025, p < .05, r = .18).

This rather small development in CT aligns with previous Anglo-American research (e.g., Arum & Roska, 2011; Astin, 1993; Bers, McGowan & Rubin, 1996; Giancarlo & Facione, 2001; Hagedorn et al., 1999; Miller, 1992; Pascarella et al., 2011; Saucier, 1995). Similar to Anglo-American higher education, Flemish higher education is successful in developing student discipline-general CT performance to a limited extent only.

8.3.2 Explanations and Ideas for Further Research

Although CT is a complex concept, a consensus on its abstract meaning was found among representatives of different institutions of higher education in Flanders (Study on CT instruments - Chapter 6). In addition, the different aspects of CT representatives agreed upon were largely represented in the different components of the HCTA and the CCTT. In the selection of the items of the Scipio these considerations were also taken into account. The conceptualisation of CT skills and dispositions as two distinct but connected factors, that are partly discipline-general is in line with the current literature on CT (Halpern, 2003; Angeli & Valenides, 2009; Ku, 2009).

Despite the agreement, translating or developing a CT test with sufficient reliability appeared to be a difficult endeavour. Limited reliability probably relates to the multi-dimensionality of CT (Cortina, 1993). Nevertheless, the strength of the correlations between the CCTT, the constructed response items and the forced-choice items of the HCTA call for additional research on the precise relation between CT skills and dispositions. It is currently unclear whether the HCTA and derived instruments are capable of independently assessing both skills and dispositions.

8.4 Research Integration and CT

8.4.1 Main Findings in Relation to Expectations

In this doctoral dissertation a clear link between research integration and CT was not found. Although students from different programmes were confronted with different research integration practices, they did not differ in their CT growth (the study on research integration and CT development – Chapter 7). This is in contrast to expectations that diverse experiences with research integration would be related to differential developments in critical thinking.

8.4.2 Explanations

A possible explanation for this finding relates to the design of the study. The study was spread over one academic year. During the first year in higher education students are confronted with a wealth of experiences (Pascarella & Tenerzini, 2005). Students in the study were simultaneously confronted with a variety of RI-approaches and stimulated to develop diverse RR-learning outcomes. The programmes studied differed from one another in multiple respects and hence students experiences differed on various dimensions. The effect of one aspect of research integration prominent in one programme can be counterbalanced by the effect of another one prominent in another programme. This limitation is intrinsically related to the ecological nature of this study and adequately mirrors reality in higher education. In addition, data included only four programmes. This makes the established differences in research integration more easily prone to coincidence.

Moreover, research integration is only one way of looking at the total set of experiences students are confronted with. Other experiences such as collaborative work, writing assignments, classroom discussions or teacher modelling (Astin, 1991; Halpern, 1998; Hofer, 2004; Pascarella & Terenzini, 2005; Tsui, 2002) could equally well influence CT development. And even more so, out of class experiences and entrance characteristics have been argued to affect critical thinking development as well (Terenzini, Springer, Pascarella & Nora, 1995; Evens, Verburgh, & Elen, 2013).

An alternative explanation for the lack of differences between programmes in development of CT, is the sensitivity of the Scipio. The Scipio may not have been sufficiently sensitive to adequately capture student development in CT. A review of Tiruneh et al. (2013) revealed that in some studies using different instruments to measure CT development, significant growth was retrieved with one instrument but not with another. Moreover, students may have mainly developed discipline-specific CT skills that are not assessed in a discipline-general test such as the Scipio. Research indicates discrepancies between development in students' discipline-specific CT and discipline-general CT (Anderson, Howe, Soden, Halliday, & Low, 2001; Renaud & Murray, 2008; Stark, 2012; Williams, Oliver, & Stockdale, 2004). Students show more development on discipline-specific CT measures. Transfer to discipline-general CT appears not self-evident (William et al., 2004). A comparison of discipline-specific CT development and discipline-general CT development could add to our understanding of the relation between research integration and CT.

The study clearly shows that instruction in higher education is not specifically designed towards CT. The claim of Browne and Freeman (2000) that CT is more common as a learning outcome than as an actual encouragement in the classroom, gets ample support. Nicholas (2011) concludes that teachers use a "Hopeful pedagogy" towards critical thinking. Teachers indeed want their students to develop CT. Teachers expect their approaches to be helpful in that respect but at the same time, they are not convinced that they actually use a good approach or that they really help students to develop their CT (Nicholas, 2011). This could also apply to teachers in this study. Opportunities for CT development seem not always seized. Nevertheless, carefully designed learning environments can induce high gains in CT (Niu, Behar-Horenstein, & Garvan, 2013; Pascarella & Terenzini, 2005). A lack of systematic design possibly results in missed chances for CT development. It is likely that the potential for effective research integration to induce CT is not exploited.

Although the study on the relationship between research integration and the development of CT (Chapter 7) was done in university colleges, there are few sound reasons to assume that the relation between research integration and critical thinking development would be largely different in university settings. The retrieved growth in CT among university students was also small and depended on the instrument (The study on CT instruments– Chapter 6), although it can be assumed that also in this context students are frequently confronted with research.

8.4.3 Implications and Ideas for Further Research

If research integration is considered important and if more research integration in the undergraduate curriculum is desired, there is a need both for more clarity with respect to research integration and for insight in the impact of learning environments with research integration. The ultimate rationale and the underlying learning outcomes for research integration are to be made explicit. Once again the importance of precise language is apparent.

To acquire the intended insights, an ambitious research agenda can be set. As a first step genuinely successful practices of research integration could be identified with controlled pre-test post-test studies. Controlled studies are necessary to avoid falsely attributing too much effects of research integration because of confounding factors (Carter et al., 2013; Seymour et al., 2004).

In a next step observation studies of the identified successful and unsuccessful practices are needed to identify key elements that possibly contribute to the development of aspired learning outcomes. This is important because research integration as such is not sufficient, as this doctoral dissertation revealed. A variety of aspects in the learning environment influence the effect of research integration.

Studies point for example at the crucial role of the supervisor in the effect of students' research experiences (e.g., Samarapungavan et al., 2006; Thiry & Laursen, 2011).

In the last step design studies could reveal which of the elements identified in the previous step indeed contribute to desired student learning. The studies could make use of the CT literature on successful approaches to foster CT. This would bring two separate lines of research fruitfully together.

8.5 Conclusion

This doctoral dissertation sheds light on particular aspects of research-integration. It shows that research integration, in terms of RR-learning outcomes is, at least to some extent, part of students' daily experiences. The same counts for RI-approaches. Clear evidence that differences with respect to research integration result in differences with respect to CT development could not be presented.

This doctoral dissertation points out that a fine-grained analysis of research integration practices at the right level of analysis adds to the understanding of the complexities involved. The findings illustrate the necessity of avoiding vague or general phrasing in favour of a very precise use of language when discussing or studying research integration (Trowler & Wareham, 2008, Spronken-Smith et al., 2012). Otherwise essential differences will remain to be missed.

The literature on research integration needs a detailed and manageable research model in order to accurately describe research integration practices and for studying effects of differences in research integration. It would help researchers to explicitly specify what is studied and develop rigorous research designs. It would help programme directors and teachers as well as policy makers to make informed decisions and to enhance the quality of learning environments. The clearly specified

RR-learning outcomes and RI-approaches could contribute to the establishment of such a research model.

This doctoral dissertation shows that research integration is not a panacea. The results exemplify the warning for falsely assuming simplistic relations between research and teaching/learning (Hattie & Marsh, 1996; Terenzini & Pascarella, 1994). The study contends the claim that research integration in the undergraduate curriculum leads to CT development and the related claim that more research should be integrated in the curriculum. The mere fact that research is integrated in the curriculum is insufficient to induce a large development of CT. If research integration as well as the development of CT are desired, much more research is to be done. Such educational research would help to improve the effectiveness of the integration of disciplinary research into teaching for the benefit of student learning

References

Abrami, P. C., Bernard, R. M., Borokhovski, E., Wade, A., Surkes, M., Tamin, R., & Zhang, D. (2008). Instructional interventions affecting critical thinking skills and dispositions: A stage 1 meta-analysis. *Review of Educational Research, 78,* 1102-1134. doi:10.3102/0034654308326084

Adhikari, N., & Nolan, D. (2002). "But what good came of it at last?": How to assess the value of undergraduate research. *Notices of the American Mathematical Society, 49,* 1252-1257.

Allen, M. (1996). Research productivity and positive teaching evaluations: Examining the relationship using meta-analysis. *Journal of the Association for Communication Administration (JACA),* 77-96.

American College Testing Program (1991). *CAAP technical handbook*. Iowa City, IA: American College Testing Program.

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching and assessing. A revision of Bloom's taxonomy of educational objectives*. New York: Longman.

Anderson, T., Howe, C., Soden, R., Halliday, J., & Low, J. (2001). Peer interaction and the learning of critical thinking skills in further education students. *Instructional Science, 29,* 1-32. doi:10.1023/A:1026471702353

Angeli, C., & Valanides, N. (2009). Instructional effects on critical thinking: Performance on ill-defined issues. *Learning and Instruction, 19,* 322-334. doi:10.1016/j.learninstruc.2008.06.010

Annala, J., & Mäkinen, M. (2011). The research-teaching nexus in higher education curriculum design. *Transnational Curriculum Inquiry, 8,* 3-21.

Arum, R., & Roksa, J. (2011). *Academically adrift: Limited learning on college campuses*. Chicago: The university of Chicago Press.

Astin, A. W. (1993). *What matters in college? Four critical years revisited*. San Fransico: Jossey Bass.

Bailin, S., Case, R., Coombs, J. R., & Daniels, L. B. (1999). Conceptualizing critical thinking. *Journal of Curriculum Studies, 31,* 285-302. doi:10.1080/002202799183133

Barnett, R. (1990). *The idea of higher education*. Buckingham: Open University Press.

Barnett, R. (2000). *Realizing the university in an age of super-complexity*. Buckingham: Open University Press.

Barnett, R. (2005). *Reshaping the university. New relationships between research, scholarship and teaching*. Berkshire, England: The society for research into higher education & Open University Press.

Bauer, K. W., & Bennett, J. S. (2003). Alumni perception used to assess undergraduate research experience. *Journal of Higher Education, 74,* 210-230. doi:10.1353/jhe.2003.0011

Baxter Magolda, M. B. (2004). Evolution of a constructivist conceptualization of epistemological reflection. *Educational Psychologist, 39,* 31-42. doi:10.1207/s15326985ep3901_4

Becher, T., & Kogan, M. (1992). *Process and structure in higher education* (2nd ed.). Routledge: London, UK.

Beckman, M., & Hensel, N. (2009). Making explicit the implicit: Defining undergraduate research. *CUR Quarterly, 29,* 40-44.

Bers, T. H., McGowan, M., & Rubin, A. (1996). The disposition to think critically among community college students: The california critical thinking dispositions inventory. *The Journal of General Education, 45,* 197-223.

Biglan, A. (1973). The characteristics of subject matter in different academic areas. *Journal of Applied Psychology, 57,* 195-203. doi:10.1037/h0034701

Bills, D. (2004). Supervisors' conceptions of research and the implications for supervisor development. *International Journal for Academic Development, 9,* 85-97. doi:10.1080/1360144042000296099

Bleiklie, I. (2003). Hierarchy and specialisation: On the institutional integration of higher education systems. *European Journal of Education, 38,* 341-355. doi:10.1111/j.0141-8211.2003.00153.x

Bock, R. D., Gibbons, R., & Muraki, E. (1988). Full-Information Item Factor Analysis. *Applied Psychological Measurement, 12,* 261-280. doi:10.1177/014662168801200305

Bok, D. (2006). *Our underachieving colleges*. Princeton: Princeton University Press.

Boyer Commission (1998). *Reinventing undergraduate education: A blueprint for America's research universities*. San Francisco: Jossey-Bass.

Brabeck, M. M. (1983). The relationship beteen critical thinking skills and the development of reflective judgement: Redefining the aims of higher education. *Journal of Applied Developmental Psychology, 4,* 23-34. doi:10.1016/0193-3973(83)90056-4

Braxton, J. M. (1996). Contrasting perspectives on the relationship between teaching and research. *New Directions for Institutional Research, 90,* 5-14. doi:10.1002/ir.37019969003

Brew, A. (1999). Research and teaching: Changing relationships in a changing context. *Studies in Higher Education, 24,* 291-301. doi:10.1080/03075079912331379905

Brew, A. (2001). Conceptions of research: A phenomenographic study. *Studies in Higher Education, 26,* 272-285. doi:10.1080/03075070120076255

Brew, A. (2003). Teaching and research: New relationships and their implications for inquiry-based teaching and learning in higher education. *Higher Education Research & Development, 22,* 3-18. doi:10.1080/0729436032000056571

Brew, A. (2006). *Research and teaching: Beyond the divide*. Hampshire: Palgrave Macmillan.

Brew, A. (2010). An Australian perspective on undergraduate research. *CUR Quarterly, 31,* 37-42.

Brew, A. (2013). Understanding the scope of undergraduate research: a framework for curricular and pedagogical decision-making. *Higher Education,* 1-16. doi:10.1007/s10734-013-9624-x

Brew, A., & Boud, D. (1995). Teaching and research: Establishing the vital link with learning. *Higher Education, 29,* 261-273. doi:10.1007/BF01384493

Brown, R. B., & Mc Cartney, S. (1998). The link between research and teaching: Its purpose and implications. *Innovations in Education and Training International, 35,* 117-129. doi:10.1080/1355800980350205

Browne, M. N., & Freeman, K. (2000). Distinguishing features of critical thinking classrooms. *Teaching in Higher Education, 5,* 301-309. doi:10.1080/713699143

Burgess, T. (1973). *The shape of higher education*. London: Cornmarket Press.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. New York: Springer.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research, 33,* 261-304. doi:10.1177/0049124104268644

Butler, H. (2012). Halpern Critical Thinking Assessment predicts real-world outcomes of critical thinking. *Applied Cognitive Psychology, 26,* 721-729. doi:10.1002/acp.2851

Butler, H., Dwyer, C., Hogan, M., Franco, A., Rivas, S., Saiz, C., & Almeida, L. (2012). The Halpern Critical Thinking Assessment and real-world outcomes: Cross national applications. *Thinking Skills and Creativity, 7,* 112-121. doi:10.1016/j.tsc.2012.04.001

Carter, D., Ro, H., Alcott, B., & Lattuca, L. R. (2013, April). *Co-curricular connections: The role of undergraduate research experiences in promoting engineering students' communications, teamwork and leadership skills*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Casar, J. R. (2000). Encouraging students' attitude of innovation in research universities. *European Journal of Engineering Education, 25,* 115-121. doi:10.1080/030437900308508

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of statistical software, 48,* 1-29.

Chan, N.-M., Ho, I. T., & Ku, K. Y. L. (2011). Epistemic beliefs and critical thinking of Chinese students. *Learning and Individual Differences, 21,* 67-77. doi:10.1016/j.lindif.2010.11.001

Clark, B. R. (1997). The modern integration of research activities with teaching and learning. *Journal of Higher Education, 68,* 241-255. doi:10.2307/2960040

Clifford, J. S., Boufal, M. M., & Kurtz, J. E. (2004). Personality traits and critical thinking skills in college students: Empirical tests of a two-factor theory. *Assessment, 11,* 169-176. doi:10.1177/1073191104263250

Coate, K., Barnett, R., & Williams, G. (2001). Relationships between teaching and research in higher education in England. *Higher Education Quarterly, 55,* 158-174. doi:10.1111/1468-2273.00180

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, New Jersey: Lawrence Erlbaum.

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th ed.). London: Routledge.

Colbeck, C. L. (1998). Merging in a seamless blend: How faculty integrate teaching and research. *Journal of Higher Education, 69,* 647-671.

Cook, P., Johnson, R., Moore, P., Myers, P., Pauly, S., Pendarvis, F. et al. (1996). *Critical Thinking assessment: Measuring a moving target. Report & recommendations of the South Carolina Higher Education Assessment Network Critical Thinking Task Force*. Rock Hill, SC: The South Carolina Higher Education Assessment Network.

Cortina, J. M. (1993). What is coefficient Alpha? An examination of theory and applications. *Journal of Applied Psychology, 78,* 98-104. doi:10.1037/0021-9010.78.1.98

Craney, C., McKay, T., Mazzeo, A., Morris, J., Prigodich, C., & de Groot, R. (2011). Cross-discipline perceptions of the undergraduate research experience. *The Journal of Higher Education, 82,* 92-113. doi:10.1353/jhe.2011.0000

Davies, M. (2011). Introduction to the special issue on critical thinking in higher education. *Higher Education Research & Development, 30,* 255-260. doi:10.1080/07294360.2011.562145

De Wit, K., & Verhoeven, J. C. (2003). The Context Changes but the Divisions Remain: The binary higher education system in Flanders. The case of information science. *Studies in Higher Education, 28,* 143. doi:10.1080/0307507032000058145

Durning, B., & Jenkins, A. (2005). Teaching/research relations in departments: The perspectives of built environment academics. *Studies in Higher Education, 30,* 407-426. doi:10.1080/03075070500160046

Ebel, R., & Frisbie, D. (1991). *Essentials of educational measurement*. Englewood Cliffs, NJ: Prentice Hall.

Education and examination regulations at the KU Leuven academic year 2011-2012, $44 and $62 (2011). Retrieved April 28, 2011, from http://www.kuleuven.be/education/regulations/

Efklides, A. (2008). Metacognition: Defining Its Facets and Levels of Functioning in Relation to Self-Regulation and Co-regulation. *European Psychologist, 13,* 277-287. doi:10.1027/1016-9040.13.4.277

Elen, J., Schouteden, W., Verburgh, A., & François, S. (2009). *Integratie van onderzoek in onderwijs: Ambities en realisatie binnen de associatie KU Leuven (IOO-project). Inhoudelijk eindverslag van OOF-project 2006/03 [Integration of research into teaching: Ambitions and realisation in the Association KU Leuven. Content report]*. Leuven, Belgium: Associatie KU Leuven.

Elen, J., Schouteden, W., Verburgh, A., & François, S. (2011). *Integratie van onderzoek in onderwijs: Realisaties en percepties (IOO² project) [Integration of research into teaching: Realisations and perceptions]*. Leuven, Belgium: Associatie KU Leuven.

Elen, J., & Verburgh, A. (2008). *Bologna in European research-intensive universities: Implications for bachelor and master programs*. Antwerpen: Garant.

Elsen, M., Visser-Wijnveen, G. J., van der Rijst, R. M., & van Driel, J. H. (2009). How to Strengthen the Connection between Research and Teaching in Undergraduate University Education. *Higher Education Quarterly, 63,* 64-85. doi:10.1111/j.1468-2273.2008.00411.x

Elton, L. (2001). Research and teaching: Conditions for a positive link. *Teaching in Higher Education, 6,* 43-56. doi:10.1080/13562510020029590

Ennis, R. H. (1989). The degree to which critical thinking is subject specific: Clarification and needed research. *Educational Researcher, 18 (3),* 4-10. doi:10.3102/0013189X018003004

Ennis, R. H., Millman, J., & Tomko, T. N. (2005). *Cornell critical thinking test (5th ed.)*. Pacific Grove, CA: Midwest Publications.

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

Erwin, T. D. (2000). *The NPEC Sourcebook on Assessment, Volume 1: Definitions and assessment methods for critical thinking, problem solving and writing*. Washington, DC: U.S. Government Printing Office.

European University Association (2007). *Lisbon declaration. Europe's universities beyond 2010: Diversity with a common purpose*. Brussels, Belgium: EUA [Retrieved January 29, 2013 from: http://www.eua.be/fileadmin/user_upload/files/Lisbon_Convention/Lisbon_Declaration.pdf].

Evens, M., Verburgh, A., & Elen, J. (2013). *Critical thinking in college freshmen: Impact of secondary and higher education*. Manuscript submitted for publication.

Everitt, B. (1993). *Cluster Analysis* (3th ed.). London: Eward Arnold.

Facione, P. (1990). *Critical thinking: A statement of expert consensus of purposes of educational assessment and instruction*. Millbrae: California Academic Press.

Facione, P. (2010). *Critical thinking: What is it and why it counts.* Retrieved from http://www.insightassessment/pfd_files/what&why2010.pdf.

Facione, P., & Facione, N. (1992). *The california critical thinking disposition inventory*. Milbrae: California Academic Press.

Farrand-Zimbardi, K., van der Burg, N., & Myatt, P. (2010). *Undergraduate students' research experiences: Bridging the gap between teaching and research in a research-intensive university*. Brisbane: University of Queensland.

Field, A. (2009). *Discovering statistics using SPSS (and sex, drugs and rock 'n' roll)*. Newbury Park: Sage.

Gadzella, B. M., & Masten, W. G. (1998). Relation between measures of critical thinking and learning styles. *Psychological Reports, 83,* 1248-1250. doi:10.2466/pr0.1998.83.3f.1248

Gellin, A. (2003). The effect of undergraduate student involvement on critical thinking: A meta-analysis of the literature 1991-2000. *Journal of College Student Development, 44,* 746-762. doi:10.1353/csd.2003.0066

Giancarlo, C. A., & Facione, P. (2001). A look across four years at the disposition towards critical thinking among undergraduate students. *The Journal of General Education, 50,* 29-55. doi:10.1353/jge.2001.0004

Glaser, E. (1941). *An experiment in the development of critical thinking*. New York: Bureau of Publications, Teacher College, Columbia University.

Griffioen, D. M. E. (2013). *Research in higher professional education. A staff perspective.* Unpublished doctoral dissertation, University of Amsterdam, The Netherlands.

Griffioen, D. M. E., de Jong, U., & Jak, S. (2013). Research self-efficacy of lecturers in non-university higher education. *Innovations in Education and Teaching International, 50,* 25-37. doi:10.1080/14703297.2012.746512

Griffiths, R. (2004). Knowledge production and the research-teaching nexus: The case of the built environment disciplines. *Studies in Higher Education, 29,* 709-726. doi:10.1080/0307507042000287212

Gunn, V., Draper, S., & Mckendrick, M. (2008). *Research-teaching linkages: Enhancing graduate attributes: Arts, humanities and social sciences*. Mansfield: The Quality Assurance Agency for Higher Education.

Hagedorn, L. S., Pascarella, E. T., Edison, M., Braxton, J., Nora, A., & Terenzini, P. T. (1999). Institutional Context and the Development of Critical Thinking: A Research Note. *Review of Higher Education, 22,* 265-285. doi:10.1353/rhe.1999.0010

Hallett, D., Chandler, M. J., & Krettenauer, T. (2002). Disentangling the course of epistemic development: Parsing knowledge by epistemic content. *New Ideas in Psychology, 20,* 285-307. doi:S0732-118X(02)00011-9

Halpern, D. F. (1998). Teaching critical thinking for transfer across domains. *American Psychologist, 53,* 449-455. doi:10.1037//0003-066X.53.4.449

Halpern, D. F. (1999). Teaching for Critical Thinking: Helping College Students Develop the Skills and Dispositions of a Critical Thinker. *New Directions for Teaching & Learning,* 69-74. doi:10.1002/tl.8005

Halpern, D. F. (2001). Assessing the effectiveness of critical thinking instruction. *The Journal of General Education, 50,* 270-286. doi:10.1353/jge.2001.0024

Halpern, D. F. (2003). *Thought and knowledge: An introduction to critical thinking*. New Jersey: Lawrence Earlbaum.

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

Halpern, D. F. (2012). *Halpern critical thinking assessment: Manual, version 22*. Mödling: Schufried.

Harvey, L. (2005). Embedding and integrating employability. *New Directions for Institutional Research, 128,* 13-28. doi:10.1002/ir.160

Hattie, J., & Marsh, H. W. (1996). The relationship between research and teaching: A meta-analysis. *Review of Educational Research, 66,* 507-542. doi:10.3102/00346543066004507

Healey, M. (2005a). Linking research and teaching to benefit student learning. *Journal of Geography in Higher Education, 29,* 183-201. doi:10.1080/03098260500130387

Healey, M. (2005b). Linking research and teaching: Exploring disciplinary spaces and the role of inquiry-based learning. In R. Barnett (Ed.), *Reshaping the university. New relations between research, scholarship and teaching* (pp. 67-78). Berkshire: Open University Press.

Healey, M. & Jenkins, A. (2009). *Developing undergraduate research and inquiry*. York, UK: The Higher Education Academy.

Healey, M., Lannin, M., Stibbe, A., & Derounian, J. (2013). *Developing and enhancing undergraduate final year projects and dissertations*. York: Higher Education Academy (forthcoming).

Heggen, K., Karseth, B., & Kyvik, S. (2010). The relevance of research for the improvement of education and professional practice. In S. Kyvik & B. Lepori (Eds.), *The research mission of higher education institutions outside the university sector* (pp. 45-60). Dordrecht: Springer.

Heppner, P. (1988). *The problem-solving inventory manual*. Palo Alto, CA: Consulting Psychology Press.

Hofer, B. K. (2004). Exploring the dimensions of personal epistemology in differing classroom contexts: Student interpretations during the first year of college. *Contemporary Educational Psychology, 29,* 129-163. doi:10.1016/j.cedpsych.2004.01.002

Howell, D. (1992). *Statistical methods for psychology* (3th ed.). Belmont, California: Duxbury Press.

Hrycaj, P. L. (2006). An analysis of Online syllabi for credit-bearing library skills courses. *College & Research Libraries, 67,* 525-535.

Hughes, C., & Tight, M. (1995). Linking university teaching and research. *Higher Education Review, 28 (1),* 51-65.

Huisman, J. & Kaiser, F. (2001). *Fixed and fuzzy bounderies in Higher Education*. The Hague: Adviesraad voor het Wetenschaps- en Technologiebeleid.

Hunter, A. B., Laursen, S. L., & Seymour, E. (2007). Becoming a scientist: The role of undergraduate research in students' cognitive, personal, and professional development. *Science Education, 91,* 36-74. doi:10.1002/sce.20173

Integratie Decreet (2012). *Decree of 13 July 2012. Ontwerp van decreet betrefffende de integratie van de academicsche hogeschoolopleidingen in de universiteiten [Integration Decree]*. Belgisch Staatblad, 08/11/2012.

International Test Commission (2010). *International Test Commission guidelines for translating and adapting tests*. Retrieved from http://www.intestcom.org.

Jakobsson, U., & Westergren, A. (2005). Statistical methods for assessing agreement for ordinal data. *Nordic College of Caring Sciences, 19,* 427-431. doi:10.1111/j.1471-6712.2005.00368.x

Jenkins, A. (2000). The relationship between teaching and research: Where does geography stand and deliver? *Journal of Geography in Higher Education, 24,* 325-351. doi:10.1080/713677414

Jenkins, A. (2004). *A guide to the research evidence on teaching-research relations*. York: The higher education academy.

Jenkins, A., Breen, R., Lindsay, R., & Brew, A. (2003). *Reshaping teaching in higher education: Linking teaching with research*. London: Kogan Page.

Jenkins, A., Healey, M., & Zetter, R. (2007). *Linking teaching and research in disciplines and departments*. York: The Higher Education Academy.

Justice, C., Rice, J., & Warry, W. (2009). Developing useful and transferable skills: Course design to prepare students for a life of learning. *International journal for the scholarship of teaching and learning, 3,* 1-19.

Kardash, C. M. (2000). Evaluation of an Undergraduate Research Experience: Perceptions of Undergraduate Interns and Their Faculty Mentors. *Journal of Educational Psychology, 92,* 191-201. doi:10.1037//0022-0663.92.1.191

Kelly, A. V. (2009). *The curriculum: Theory and practice* (6th ed.). London: Sage.

King, P. M., & Kitchener, K. S. (1994). *Developing reflective judgement*. San Francisco, CA: Jossey Bass.

King, P. M., & Kitchener, K. S. (2002). The reflective judgement model: Twenty years of research on epistemic cognition. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 37-61). Mahwah: Lawrence Erlbaum.

King, P. M., & Kitchener, K. S. (2004). Reflective judgment: Theory and research on the development of epistemic assumptions through adulthood. *Educational Psychologist, 39,* 5-18. doi:10.1207/s15326985ep3901_2

King, P. M., Kitchener, K. S., & Wood, P. K. (2000). *The reasoning about current issues test*. Retrieved from http://www.reflectivejudgment.org/.

King, P. M., Wood, P. K., & Mines, R. A. (1990). Critical thinking among college and graduate students. *The Review of Higher Education, 13,* 167-186.

Kitchener, K. S. (2002). Skills, tasks, and definitions: Discrepancies in the understanding and data on the development of folk epistemology. *New Ideas in Psychology, 20,* 309-328. doi:10.1016/S0732-118X(02)00012-0,

Kline, P. (1999). *Handbook of psychological testing*. London: Routledge.

Kreber, C. (2000). How university teaching award winners conceptualise academic work: Some further thoughts on the meaning of scholarship. *Teaching in Higher Education, 5,* 61-78. doi:10.1080/135625100114966

Kronholm, M. (1996). The impact of developmental instruction on reflective judgement. *The Review of Higher Education, 19,* 199-225.

Ku, K. (2009). Assessing students' critical thinking performance: Urging for measures using multi-response format. *Thinking Skills and Creativity, 4,* 70-76. doi:10.1016/j.tsc.2009.02.001

Ku, K., & Ho, I. (2010). Dispositional factors predicting Chinese students' critical thinking performance. *Personality and Individual Differences, 48,* 54-58. doi:10.1016/j.paid.2009.08.015

Kuh, G. D. (2007). *Experiences that matter: Enhancing student learning and success. Annual report 2007*. Bloomington, Indiana: National Survey of student engagement.

Kuhn, D. (1999). A developmental model of critical thinking. *Educational Researcher, 28,* 16-25 +46. doi:10.3102/0013189X028002016

Kyvik, S. (2009). The transformation of the college sector. In S. Kyvik (Ed.), *The dynamics of change in higher education: Expansion and contraction* (pp. 3-20). Dordrecht: Springer. doi:10.1007/978-1-4020-9248-0_1

Kyvik, S., & Lepori, B. (2010). Research in higher education institutions outside the university sector. In S. Kyvik & B. Lepori (Eds.), *The research mission of higher education institutions outside the university sector* (pp. 3-21). Dordrecht: Springer. doi:10.1007/978-1-4020-9244-2_1

Kyvik, S., & Skodvin, O. J. (2003). Research in the non-university higher education sector - tensions and dilemmas. *Higher Education, 45,* 203-222. doi:10.1023/A:1022492218855

Kyvik, S., & Smeby, J.-C. (1994). Teaching and research. The relationship between the supervision of graduate students and faculty research performance. *Higher Education, 28,* 227-239. doi:10.1007/BF01383730

Leisyte, L., Enders, J., & de Boer, H. (2009). The balance between teaching and research in Dutch and English universities in the context of university governance reforms. *Higher Education, 58,* 619-635. doi:10.1007/s10734-009-9213-1

Lepori, B., & Kyvik, S. (2010). The research mission of universities of applied sciences and the future configuration of higher education systems in Europe. *Higher Education Policy, 23,* 295-316. doi:10.1057/hep.2010.11

Levy, P., & Petrulis, R. (2012). How do first-year university students experience inquiry and research, and what are the implications for the practice of inquiry-based learning? *Studies in Higher Education, 37,* 85-101. doi:10.1080/03075079.2010.499166

Linhart, H., & Zucchini, W. (1986). *Model selection*. New York: Wiley.

Lopatto, D. (2009). *Science in solution: The impact of undergraduate research on student learning*. Tucson Arizona: Research corporation for science advancement.

Malcolm, M. (2008). *Research-Teaching Linkages: Enhancing graduate attributes: Business, management, accountancy and finance*. Mansfield: The Quality Assurance Agency for Higher Education.

Mälkki, K., & Lindblom-Yllänne, S. (2012). From reflection to action? Barriers and bridges between higher education teachers: Thoughts and actions. *Studies in Higher Education, 37,* 33-50. doi:10.1080/03075079.2010.492500

Maneesriwongul, W., & Dixon, J. K. (2004). Instrument translation process: A methods review. *Methodological Issues in Nursing Research, 48,* 175-186. doi:10.1111/j.1365-2648.2004.03185.x

Marsh, C. J., & Willis, G. (1999). *Curriculum: Alternative approaches, ongoing issues* (2nd ed.). Upper Saddle River: Merrill.

Marsh, H. W., & Hattie, J. (2002). The relation between research productivity and teaching effectiveness: Complementary, antagonistic or independent constructs? *Journal of Higher Education, 73,* 603-641. doi:10.1353/jhe.2002.0047

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd ed.). Mahwah: Erlbaum.

McAlpine, L., Weston, C., Timmermans, J., Berthiaume, D., & Fairban-Roch, G. (1-10-2006). Zones: Reconceptualizing teacher thinking in relation to action. *Studies in Higher Education, 31,* 601-615. doi:10.1080/03075070600923426

McPeck, H. (1990). Critical thinking and subject specificity: A reply to Ennis. *Educational Researcher, 19,* 10-12. doi:10.3102/0013189X019004010

Mertens, D. (2010). *Research and evaluation in education and psychology* (3th ed.). Thousand Oaks: Sage.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis. An expanded sourcebook* (2nd ed.). Thousand Oaks: Sage.

Miller, G., & Chapman, J. (2001). Misunderstanding Analysis of Covariance. *Journal of Abnormal Psychology, 110,* 40-48. doi:10.1037//0021-843X.110.1.40

Miller, M. A. (1992). Outcomes evaluation. Measuring critical thinking. *Journal of Advanced Nursing, 17,* 1401-1407. doi:10.1111/j.1365-2648.1992.tb02810.x

Moore, T. (2004). The critical thinking debate: How general are general thinking skills? *Higher Education Research & Development, 23,* 3-18. doi:10.1080/0729436032000168469

Moore, T. (2013). Critical thinking: Seven definitions in search of a concept. *Studies in Higher Education, 38,* 506-522. doi:10.1080/03075079.2011.586995

Muchinsky, P. M. (1996). The Correction for Attenuation. *Educational and Psychological Measurement, 56,* 63-75. doi:10.1177/0013164496056001004

Murphy, K. R., & Davidshofer, C. O. (1991). *Psychological testing. Principles and applications* (2th ed.). Englewood Cliffs, New Jersey: Prentice-Hall International Editions.

Murray, K., & Macdonald, R. (1997). The disjunction between lecturers' conceptions of teaching and their claimed educational practice. *Higher Education, 33,* 331-349. doi:10.1023/A:1002931104852

Neuman, W. L. (2011). *Social research methods: Qualitative and quantitative approaches* (7th ed.). Boston: Pearson.

Neumann, R. (1992). Perceptions of the teaching-research nexus: A framework for analysis. *Higher Education, 23,* 159-171. doi:10.1007/BF00143643

Neumann, R. (1993). Research and Scholarship: Perceptions of Senior Academic Administrators. *Higher Education, 25,* 97-110. doi:10.1007/BF01384743

Neumann, R. (1994). The teaching research nexus: Applying a framework to university students' learning experiences. *European Journal of Education, 29,* 323-338. doi:10.2307/1503744

Nicholas, M. (2011). *Faculty conceptualizations and approaches to assessing critical thinking in the humanities and natural sciences - A grounded theory study.* (Electronic Thesis or Dissertation). Retrieved from https://etd.ohiolink.edu/.

Nieto, A. M., Saiz, C., & Orgaz, B. (2009). Análisis de la propiedades psicométricas de la versión espanola del HCTAES-test de Halpern para la evaluaciòn del pensamiento crítico mediante situaciones cotidianas [Analysis of the psychometric properties of the Spanish version of Halpern HCTAES-test for the assessment of critical thinking through everyday situations]. *Revista Electrónica de Metodología Apliciada, 14,* 1-15.

Nisbet, J. D., & Entwistle, N. J. (1970). *Educational research methods*. London: University of London Press.

Niu, L., Behar-Horenstein, L. S., & Garvan, C. W. (2013). Do instructional interventions influence college students critical thinking skills? A meta-analysis. *Educational Research Review, 9,* 114-128. doi:10.1016/j.edurev.2012.12.002

OECD (1998). *Redefining tertiary education*. Paris: OECD.

OECD (2013). *Assessment of higher education learning outcomes AHELO: Feasibility study resport*. Paris: OECD.

Parkes, J., & Harris, M. B. (2002). The Purposes of a Syllabus. *College Teaching, 50,* 1. doi:10.1080/87567550209595875

Pascarella, E. T., Blaich, C., Martin, G. L., & Hanson, J. M. (2011). How Robust Are the Findings of Academically Adrift? *Change, 43,* 20-24. doi:10.1080/00091383.2011.568898

Pascarella, E. T., & Terenzini, P. T. (1991). *How college affects students: Findings and insight from twenty years of reserarch*. San Francisco: Jossey-Bass.

Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students. Volume 2: A third decade of research*. San Francisco: Jossey-Bass.

Paul, R., Elder, L., & Bartell, T. (1997). *Study of 38 Public Universities and 28 Private Universities to determine faculty emphasis on critical thinking in instruction. Executive summary.* Sacramento, California: California Commission on Teacher Credentialing.

Perry, W. G. (1970). *Forms of intellectual and ethical development in the college years*. New York: Holt, Renehart and Winston.

Pithers, R. T., & Soden, R. (2000). Critical thinking in education: A review. *Educational Research, 42,* 237-249. doi:10.1080/001318800440579

Prince, M. J., Felder, R. M., & Brent, R. (2007). Does faculty research improve undergraduate teaching? An analysis of existing and potential synergies. *Journal of Engineering Education, 96,* 282-294. doi:10.1002/j.2168-9830.2007.tb00939.x

Prokou, E. (2008). The emphasis on employability and the changing role of the university in Europe. *Higher Education in Europe, 33,* 387-394. doi:10.1080/03797720802522593

Prosser, M., Martin, E., Trigwell, K., Ramsden, P., & Lueckenhausen, G. (2005). Academics' experiences of understanding of their subject matter and the relationship of this to their experiences of teaching and learning. *Instructional Science, 33,* 137-157. doi:10.1007/s11251-004-7687-x

Rauckhorst, W. H., Czaja, J. J. A., & Baxter Magolda, M. B. (2001, July). *Measuring the impact of the undergraduate research experience on student intellectual development*. Paper presented at Project Kaleidoscope Summer Institute, Snowbird, UT.

Reason, R., Terenzini, P., & Domingo, R. (2006). First Things First: Developing Academic Competence in the First Year of College. *Research in Higher Education, 47,* 149-175. doi:10.1007/s11162-005-8884-4

Reisman, A. (2013). *Developing text-based student argumentation in whole-class discussions about the past*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Renaud, R. D., & Murray, H. G. (2008). A comparison of a subject-specific and a general measure of critical thinking. *Thinking Skills and Creativity, 3,* 85-93. doi:10.1016/j.tsc.2008.03.005

Rhodes, F. H. T. (2004). Reinventing the university. In L. E. Weber & J. J. Duderstadt (Eds.), *Reinventing the university* (pp. 3-13). London: Economica.

Robertson, J. (2007). Beyond the 'research/teaching nexus': Exploring the complexity of academic experience. *Studies in Higher Education, 32,* 541-556. doi:10.1080/03075070701476043

Robertson, J., & Bond, C. H. (2001). Experiences of the relation between teaching and research: What do academics value? *Higher Education Research & Development, 20,* 5-19. doi:10.1080/07924360120043612

Rowland, S. (1996). Relationships between teaching and research. *Teaching in Higher Education, 1,* 7-20. doi:10.1080/1356251960010102

Ryder, J., Leach, J., & Driver, R. (1999). Undergraduate science students' images of science. *Journal of Research in Science Teaching, 36,* 201-219. doi:10.1002/(SICI)1098-2736(199902)36:2<201::AID-TEA6>3.0.CO;2-H

Saiz, C., & Rivas, S. (2008). Assessment in critical thinking: A proposal for differentiating ways of thinking. *Ergo Nueva Epoca, 22-23,* 25-66.

Samarapungavan, A., Westby, E. L., & Bodner, G. M. (2006). Contextual epistemic development in science: A comparison of chemistry students and research chemists. *Science Education, 90,* 468-495. doi:10.1002/sce.20111

Saucier, B. L. (1995). Critical thinking skills of baccalaureate nursing students. *Journal of Professional Nursing, 11,* 351-357. doi:10.1016/S8755-7223(95)80054-9

Scheerens, J., Glas, C., & Thomas, S. M. (2003). *Educational evaluation, assessment and monitoring: A systematic approach*. Lisse: Swets & Zeitlinger.

Schommer, M., & Walker, K. (1995). Are epistemological beliefs similar across domains? *Journal of Educational Psychology, 87,* 424-432. doi:10.1037//0022-0663.87.3.424

Schouteden, W., Verburgh, A., & Elen, J. (2012, September). *Lecturers in professional fields in higher education: Conception of research and it perceived importance of research skills for students' careers.* Paper presented at the annual conference of ECER, Cadiz, Spain.

Schouteden, W., Verburgh, A., & Elen, J. (2013). *Teachers' research conceptions educationally interpreted*. Manuscript submitted for publication.

Seifert, T. A., Pascarella, E. T., Colangelo, N., & Assouline, S. (2007). The effects of honors program participation on experiences of good practices and learning outcomes. *Journal of College Student Development, 48,* 57-74. doi:10.1353/csd.2007.0007

Seymour, E., Hunter, A. B., Laursen, S. L., & Deantoni, T. (2004). Establishing the benefits of research experiences for undergraduates in the sciences: First findings from a three-year study. *Science Education,* 493-534. doi:10.1002/sce.10131

Simons, M. (2006). Education Through Research at European Universities: Notes on the Orientation of Academic Research. *Journal of Philosophy of Education, 40,* 31-50. doi:10.1111/j.1467-9752.2006.00493.x

Simons, M., & Elen, J. (2007). The 'research-teaching nexus' and 'education through research': An exploration of ambivalences. *Studies in Higher Education, 32,* 617-631. doi:10.1080/03075070701573781

Sin, C. (2012). Researching research in master's degrees in Europe. *European Educational Research Journal, 11,* 290-301. doi:10.2304/eerj.2012.11.2.290

Spronken-Smith, R., & Walker, R. (2010). Can inquiry-based learning strengthen the links between teaching and disciplinary research? *Studies in Higher Education, 35,* 723-740. doi:10.1080/03075070903315502

Spronken-Smith, R., Walker, R., Batchelor, J., O'Steen, B., & Angelo, T. (2012). Evaluating student perceptions of learning processes and intended learning outcomes under inquiry approaches. *Assessment & Evaluation in Higher Education, 37,* 57-72. doi:10.1080/02602938.2010.496531

Stark, E. (2012). Enhancing and assessing critical thinking in a psychological research methods course. *Teaching of Psychology, 39,* 107-112. doi:10.1177/0098628312437725

Stark, J. S., Briggs, C. L., & Rowland-Poplawski, J. (2002). Curriculum leadership roles of chairpersons in continuously planning departements. *Research in Higher Education, 43,* 329-356. doi:10.1023/A:1014841118080

Stark, J. S., & Lattuca, L. R. (1997). *Shaping the college curriculum. Academic plans in action*. Boston: Allyn and Bacon.

Tapper, J. (2004). Student perceptions of how critical thinking is embedded in a degree program. *Higher Education Research & Development, 23,* 199-222. doi:10.1080/0729436042000206663

Taylor, J. (2007). The teaching-research nexus: A model for institutional management. *Higher Education, 54,* 867-884. doi:10.1007/s10734-006-9029-1

Terenzini, P. T., & Pascarella, E. T. (1994). Living with Myths: Undergraduate Education in America. *Change, 26,* 28-32. doi:10.1080/00091383.1994.9938488

Terenzini, P. T., & Reason, R. (2005, November). *Parsing the first year of college: A conceptual framework for studying college impacts*. Paper presented at the meeting of the Association for Study of Higher Education, Philadelphia, PA.

Terenzini, P. T., & Reason, R. (in press). Rethinking between-college effects on student learning: A new model to guide assessment and quality assurance. In R. Yamada (Ed.), *Measuring quality of undergraduate education in Japan: Comparative perspective of quality assurance* (Dordrecht: Springer.

Terenzini, P. T., Ro, H., & Yin, A. (2010, November). *Between-college effects on students reconsidered.* Paper presented to the meeting of the Association for the Study of Higher Education, Indianapolis, IN .

Terenzini, P. T., Springer, L., Pascarella, E. T., & Nora, A. (1995). Influences Affecting the Development of Students Critical Thinking Skills. *Research in Higher Education, 36,* 23-39. doi:10.1007/BF02207765

Terenzini, P. T., Springer, L., Yaeger, P. M., Pascarella, E. T., & Nora, A. (1996). First-generation college students: Characteristics, experiences, and cognitive development. *Research in Higher Education, 37,* 1-22. doi:10.1007/BF01680039

Thiry, H., & Laursen, S. (2011). The Role of Student-Advisor Interactions in Apprenticing Undergraduate Researchers into a Scientific Community of Practice. *Journal of Science Education and Technology, 20,* 771-784. doi:10.1007/s10956-010-9271-2

Tiruneh, D., Verburgh, A., & Elen, J. (2013). *Effectiveness of critical thinking instruction in higher education: A review of the intervention studies*. Manuscript under review.

Trowler, P. & Wareham, T. (2008). *Tribes, territories, research and teaching. Enhancing the teaching-research nexus*. York: The higher education Academy.

Tsui, L. (1999). Courses and instruction affecting critical thinking. *Research in Higher Education, 40,* 185-200. doi:10.1023/A:1018734630124

Tsui, L. (2001). Faculty attitudies and the development of students' critical thinking. *The Journal of General Education, 50,* 1-28. doi:10.1353/jge.2001.0008

Tsui, L. (2002). Fostering critical thinking through effective pedagogy: Evidence from four institutional cases. *The Journal of Higher Education, 73,* 740-763. doi:10.1353/jhe.2002.0056

Valanides, N., & Angeli, C. (2005). Effects of instruction on changes in epistemological beliefs. *Contemporary Educational Psychology, 30,* 314-330. doi:10.1016/j.cedpsych.2005.01.001

Van den Akker, J. (2003). Curriculum perspectives: An introduction. In J. Van den Akker, W. Kuiper, & U. Hameyer (Eds.), *Curriculum landscapes and trends* (pp. 1-10). Dordrecht: Kluwer Academic Press.

van der Rijst, R. M. (2009). *The research-teaching nexus in the sciences. Scientific research dispositions and teaching practice*. Unpublished doctoral dissertation. University of Leiden, Leiden.

van der Rijst, R. M. & Jacobi, R. (2009). *Verwevenheid van onderzoek in het onderwijs aan de Universiteit Leiden: Curriculumanalyses van bacheloropleidingen archeologie, MST en criminologie [Integration of research into teaching at Leiden University: Analyses of the bachelor curricula of Archeology, MST and Criminology]* (Rep. No. 191). Leiden: ICLON.

van der Rijst, R. M., & Jacobi, R. (2010). De verwevenheid van onderzoek en onderwijs in universitaire bacheloropleidingen [The teaching-research nexus in bachelor programmes at the university]. *Tijdschrift voor Hoger Onderwijs,* 1-20.

Van Hertbruggen, C. (2013). *Onderzoeksopvattingen onderwijskundig bekeken [Research conceptions for an educational point of view]*. (Unpublished master's thesis). KU Leuven, Faculteit Psychologie en Pedagogische wetenschappen, Leuven, Belgium.

Vandermensbrugghe, J. (2004). The unbearable vagueness of critical thinking in the context of the verhoeven 2008Questioning the binary divide: Non-university higher education in Flanders (Belgium). 2008Anglo-Saxonisation of Education. *International Education Journal, 5,* 417-422.

Verburgh, A., Elen, J., & Lindblom-Ylänne, S. (2007). Investigating the myth of the relation between teaching and research in higher education: A review of empirical research. *Studies in Philosophy and Education, 26,* 449-465. doi:10.1007/s11217-007-9055-1

Verburgh, A., François, S., & Elen, J. (2012, September). *The development of critical thinking in first year students*. Paper presented at the annual conference of ECER, Cádiz, Spain.

Verburgh, A., Schouteden, W., & Elen, J. (2009, August). *A look at the prevalence of research-related educational goals in educational programmes of higher education*. Paper presented at the bi-annual meeting of EARLI, Amsterdam, The Netherlands.

Verburgh, A., Schouteden, W., & Elen, J. (2013a). *Research integration in the first year: Relationships between research-related learning outcomes and research integration approaches* . Manuscript submitted for publication.

Verburgh, A., Schouteden, W., & Elen, J. (2013b). *The prevalence of research integration approaches in the undergraduate curriculum*. Manuscript submitted for publication.

Verburgh, A. L., Schouteden, W., & Elen, J. (2013c). Patterns in the prevalence of research-related goals in higher education programmes. *Teaching in Higher Education, 18,* 298-310. doi:10.1080/13562517.2012.719153

Verburgh, A., Schouteden, W., Evens, M., Tiruneh, D., & Elen, J. (2013). *Development of critical thinking in the first year of higher education: The influence of research integration.* Manuscript under review.

Verhoeven, J. (2008). Questioning the binary divide: Non-university higher education in Flanders (Belgium). In J. Taylor, J. Brites Fereira, M. Machado, & R. Santiago (Eds.), *Non-University higher education in Europe. Higher education dynamics* (pp. 43-75).

Verhoeven, J. C. (2010). Research in university colleges in Belgium. In S. Kyvik & B. Lepori (Eds.), *The research mission of higher education institutions outside the university sector* (pp. 97-114). New York: Springer. doi:10.1007/978-1-4020-9244-2_6

Verhoeven, J. C. & Beuselinck, I. (1996). *Higher Education in Flanders (Belgium). A report for the OECD* Departement of Sociology, Centre for Theoretical Sociology and Sociology of Education.

Visser-Wijnveen, G. J. (2009). *The teaching-research nexus in the humanities. Variations among academics*. Unpublished doctoral dissertation. University of Leiden, Leiden.

Visser-Wijnveen, G. J., van Driel, J. H., van der Rijst, R. M., Verloop, N., & Visser, A. (2010). The ideal research-teaching nexus in the eyes of academics: Building profiles. *Higher Education Research & Development, 29,* 195-210. doi:10.1080/07294360903532016

Visser-Wijnveen, G. J., van Driel, J. H., van der Rijst, R. M., Visser, A., & Verloop, N. (2012). Relating academics' ways of integrating research and teaching to their students' perceptions. *Studies in Higher Education, 37,* 219-234. doi:10.1080/03075079.2010.536913

VLIR-VLHORA (2008). *Handleiding onderwijsvisitaties, Deel 1 [Manual Educational Visitation, Part 1]*. Brussel, Belgium: VLIR-VLHORA.

VLUHR (2013). *Concept Visitatieprotocol (Hoofdstukken 1-5) [Outline of the quality assurance protocol (Chapters 1-5)]*. Brussel, Belgium: VLUHR.

Vogt, W. P. (1999). *Dictionary of statistics and methodology: A nontechnical guide for the social sciences*. Newbury Park: Sage.

Wang, W.-L., Lee, H.-L., & Fetzer, S. J. (2006). Challenges and strategies of instrument translation. *Western Journal of Nursing Research, 28,* 310-321. doi:10.1177/0193945905284712

Watson, G., & Glaser, E. M. (1980). *Watson-Glaser critical thinking appraisal*. Cleveland, OH: Psychological Corporation.

Westergaard, J. (1991). Scholarship, Research and Teaching: A View from the Social Sciences. *Studies in Higher Education, 16,* 23-28. doi:10.1080/03075079112331383061

Williams, R. L., Oliver, R., & Stockdale, S. (2004). Psychological versus generic critical thinking as predictors and outcome measures in a large undergraduate human development course. *The Journal of General Education, 53,* 37-58. doi:10.1353/jge.2004.0022

Willison, J., & O'Regan, K. (2007). Commonly known, commonly not known, totally unknown: A framework for students becoming researchers. *Higher Education Research & Development, 26,* 393-409. doi:10.1080/07294360701658609

Witte, J., van der Wende, M., & Huisman, J. (2008). Blurring boundaries: How the Bologna process changes the relationship between university and non-university higher education in Germany, the Netherlands and France. *Studies in Higher Education, 33,* 217-231. doi:10.1080/03075070802049129

Wood, P. K., & Kardash, C. (2002). Critical elements in the design and analysis of studies of epistemology. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 231-260). Mahwah: Lawrence Erlbaum.

Wood, P. K., Kitchener, K. S., & Jensen, L. (2002). Considerations in the design and evaluation of a paper-and-pencil measure of epistemic cognition. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 277-294). Mahwah: Lawrence Erlbaum.

Yeh, S. (2001). Tests worth teaching to: Constructing state-mandated tests that emphasize critical thinking. *Educational Researcher, 30,* 12-17. doi:10.3102/0013189X030009012

Zamorski, B. (2002). Research-led teaching and learning in higher education: A case. *Teaching in Higher Education, 7,* 411-427. doi:10.1080/135625102760553919

Zimbardi, K., & Myatt, P. (2012). Embedding undergraduate research experiences within the curriculum: a cross-disciplinary study of the key characteristics guiding implementation.

Studies in Higher Education. Advance online publication. doi:10.1080/03075079.2011.651448

Zubrick, A., Reid, I., & Rossiter, P. (2001). *Strengthening the nexus between teaching and research* Canberra: Department of Education, Training and Youth Affairs. Retrieved December 12, 2005, from: http://www.dest.gov.au/archive/highered/eippubs/eip01_2/01_2.pdf.